Functional Data Models for Raman Spectral Data and Degradation Analysis

Quyen Ngoc Do

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

Pang Du, Chair
Inyoung Kim
Xinwei Deng
Yili Hong

August 3, 2022
Blacksburg, Virginia

Keywords: Functional data analysis, degradation data analysis, functional linear regression, nonparametric regression, Raman spectra, Lithium-ion batteries

Functional Data Models for Raman Spectral Data and Degradation Analysis

Quyen Ngoc Do

(ABSTRACT)

Functional data analysis (FDA) studies data in the form of measurements over a domain as whole entities. Our first focus is on the post-hoc analysis with pairwise and contrast comparisons of the popular functional ANOVA model comparing groups of functional data. Existing contrast tests assume independent functional observations within group. In reality, this assumption may not be satisfactory since functional data are often collected continually overtime on a subject. In this work, we introduce a new linear contrast test that accounts for time dependency among functional group members. For a significant contrast test, it can be beneficial to identify the region of significant difference. In the second part, we propose a non-parametric regression procedure to obtain a locally sparse estimate of functional contrast. Our work is motivated by a biomedical study using Raman spectroscopy to monitor hemodialysis treatment near real-time. With contrast test and sparse estimation, practitioners can monitor the progress of the hemodialysis within session and identify important chemicals for dialysis adequacy monitoring. In the third part, we propose a functional data model for degradation analysis of functional data. Motivated by degradation analysis application of rechargeable Li-ion batteries, we combine state-of-the-art functional linear models to produce fully functional prediction for curves on heterogenous domains. Simulation studies and data analysis demonstrate the advantage of the proposed method in predicting degradation measure than existing method using aggregation method.

Functional Data Models for Raman Spectral Data and Degradation Analysis

Quyen Ngoc Do

(GENERAL AUDIENCE ABSTRACT)

Functional data analysis (FDA) studies complex data structure in the form of curves and shapes. Our work is motivated by two applications concerning data from Raman spectroscopy and battery degradation study. Raman spectra of a liquid sample are curves with measurements over a domain of wavelengths that can identify chemical composition and whose values signify the constituent concentrations in the sample. We first propose a statistical procedure to test the significance of a functional contrast formed by spectra collected at beginning and at later time points during a dialysis session. Then a follow-up procedure is developed to produce a sparse representation of the contrast functional contrast with clearly identified zero and nonzero regions. The use of this method on contrast formed by Raman spectra of used dialysate collected at different time points during hemodialysis sessions can be adapted for evaluating the treatment efficacy in real time. In a third project, we apply state-of-the-art methodologies from FDA to a degradation study of rechargeable Li-ion batteries. Our proposed methods produce fully functional prediction of voltage discharge curves allowing flexibility in monitoring battery health.

Dedication

For Keith, Me (Mother), Ba (Father), and my sister Minh, for your unconditional love and support

Acknowledgments

I would like to thank my advisor, Dr. Pang Du for his guidance and advice over the last four years. I started working with Dr. Du after my first year at Virginia Tech and I have never ceased learning from him. He taught me how to be a researcher. I learned from him how to read, review, and write a statistics paper effectively, how to conduct research, and how to efficiently run a simulation study. He is a great teacher, mentor, and incredibly patient editor. What I learned from him in and out of the classroom helped me become a more well-rounded researcher and statistician. I would also like to thank Dr. Inyoung Kim, Dr. Xinwei Deng, and Dr. Yili Hong for graciously serving on my committee. I have taken at least two classes from each of them. I learned not only their statistical methods but their work ethic and dedication to teaching and mentoring. Dr. Kim taught me two Regression courses. I was the only student in Advanced Topics in Regression and we had to switch from in-person class to online because of the pandemic. Nonetheless, her energy and passion for the subject always inspired me. What I learned from STAT 6444 solidified my understanding of my research topic. Dr. Deng taught me two DOE courses. DOE was my weakest subject in the Qualifying Exam. However, with the teaching and guidance from Dr. Deng in DOE 2, I regained my confidence and excitement for experimental design. I'm excited to bring with me the lessons I learn from him to my future work. I always hold dear the advice he gave about life, research and statistics. I am indebted to Dr. Hong for introducing me to the project that motivated the work from Chapter 4 of my dissertation. From knowing nothing about the topic, Reliability Analysis became one of my favorite subjects of statistics thanks to his teaching and passion. He is always available for help beyond homework and research. During my graduate school journey, I'm very lucky to work with wonderful supervisors. I

want to thank Dr.Laura Sands who provided financial support for me over the last three years at Virginia Tech. It was an honor to serve as her research assistant. I had front-row seats to working with professional medical researchers, learning from their subject specialty, and contributing to research projects that help improve patient care. It was an honor to participate in her research and get to know her as a person. To Ginny Agresti, my supervisor at 3M during my summer internship. She showed me the great joy and passion of being an industrial statistician. The exciting projects Ginny assigned to me helped shape my aspiration to become a statistician, working not just for a private institution but also for the public good. The opportunity to work with and attend meetings with a master statistical communicator like Ginny showed me the gold standard of statistical collaboration. Thanks to all professors and colleagues at Virginia Tech for making this journey so exciting. Thank you, Dr. Vining for sponsoring my trip to FTC 2019, Dr. Gramacy for the challenging and fun Surrogate Modeling and Advanced Statistical Computing courses. Thank you, Dr. Franck for his support in my NSF Fellow application and his advice on writing, Dr. Ranganathan for serving on my Master's degree committee, Dr. House, Dr. Ferreira, and Dr. Leman for teaching me about Bayes. Last but not least, Dr. Van Mullekom provided me with huge support during my job search and application. I will try my best to honor the lessons on effective statistics communication I learned from her. I want to thank my cohort and friends, Ryan, Jake, Tessa, Annie, Stephen, and Kevin for all the fun and laughs in our little first-year office. Thanks to my predecessors, Sierra, and Mohamed, for their check-ins and encouragement. Thanks to Wenyu for the long conversations on school, work, and life. I want thank Adeline, Soo, Shuangshuang, and Jie for their help with homework and class projects. I cherish all the fun memories from all the hang-outs and get-togethers with Adeline, Chris, Ryan, Marissa, Young, John, Yanran, Qing, and Yueyao. Lastly, thanks to my friends at the departments and University, Betty and Saurabh for making me feel home during my time here at Virginia Tech. Thanks to Dr. Kevin Robinson, Dr. Dorothee Blum

from Millersville University and Dr. Trent Gaugler at Lafayette College for inspiring me to embark on the journey to graduate school. To my friends from Vietnam, Hong Van and Minh Anh, thank you for always believing in me and our across-continent conversations that felt like we have never been apart. To my family, Uncle Trang, Minh, Mom, Dad, my in-laws Kate and John for their undying support. Nothing lifts my spirits more than our phone calls and coming home to you. Last but not least, to Keith, the love of my life. You left your hometown and moved to Blacksburg so that we can be together while I finish my graduate work. Being married to you and building our home together over the last two years are the things I treasure most.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

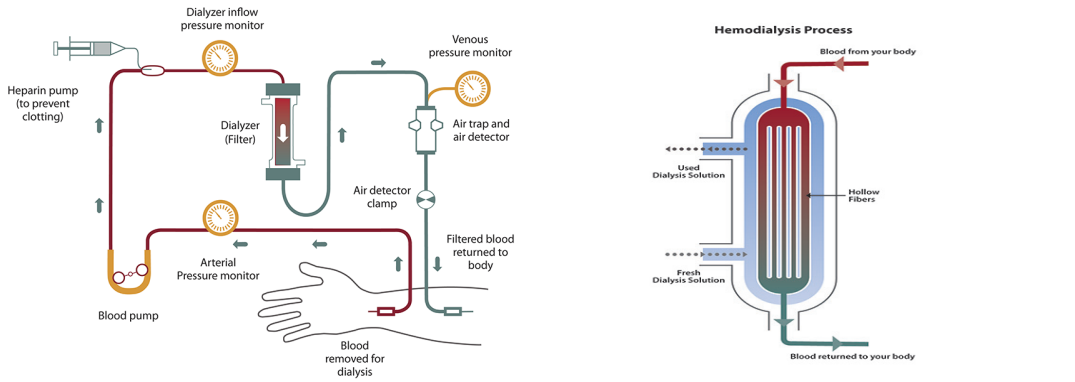
1.1 Motivating Applications

1.1.1 Monitoring hemodialysis procedure through Raman spectra data

The first two parts of this dissertation are motivated by the need to monitor the efficacy of hemodialysis procedure. Hemodialysis is one of the two types of dialysis treatment for patients with end-stage renal disease (ESRD). ESRD is caused by kidney failure and commonly treated by either kidney transplant or dialysis. Because the number of people on the kidney transplant list far exceeds the number of available kidney donors, kidney transplant is not a viable treatment for everyone and dialysis is the popular alternative treatment for ESRD patients who continue to live with kidney failure or are waiting for a kidney transplant. According to Centers for Disease Control and Prevention [8], 71% ESRD patients in the US are on dialysis. According to the Annual Data Report by the United States Renal Data System [67], the number of prevalent ESRD has steadily increased since 2000.

There are two common dialysis procedures: hemodialysis and peritoneal dialysis. As the dominant one of the two procedures, hemodialysis aims to perform the function of a kidney in filtering chemical and toxic waste out of patient's blood. During the hemodialysis procedure,

patient blood is transmitted to a dialyzer where blood is cleaned by passing through hollow fibers as dialysis solution (or dialysate) moves in opposite direction outside the fibers to collect chemical waste (Figure 1.1b). Used dialysate is removed from the dialyzer while fresh solution is continuously pumped into the dialyzer; at the same time, filtered blood exits and is transmitted back into patient body completing the filtering cycle (Figure 1.1a).



(a) Hemodialysis procedure diagram (b) Dialyzer diagram

Figure 1.1: Hemodialysis procedure [39]

It is well-known that hemodialysis is a costly and time-consuming procedure, and above all, painful experience for patients. On top of it, the filtration effect of hemodialysis decays within days requiring patients to undergo multiple treatment sessions a week with each lasts from three to four hours. Currently, the monitoring of hemodialysis for each patient is still done through monthly lab tests to obtain measures such as urea reduction ratio (URR) or kt/V. Both measures communicate the reduction in urea in patient blood post-dialysis; however, they are only useful in signifying the dialysis efficacy to the extent of the frequency of the blood work. Thus, there is a need for real-time monitoring of the dialysis procedure.

Used dialysate leaving dialyzer after filtering patient blood carries chemical compounds that provide critical information for the determination of hemodialysis efficacy in real time. In a study, Virginia Tech researchers utilize an established chemical concept called Raman spectroscopy to extract Raman spectra from liquid sample. Data from Raman spectra contain

important qualitative and quantitative information on the compounds that make up the liquid sample [60]. The biomarker ability of Raman spectra allows one to view peaks in the spectra as the intensity of a chemical component with the specific chemical component identified by the wavelength they reside in. Raman spectra can be obtained fast and non-destructively from a liquid sample. This makes it ideal for Raman spectra to be obtained from used dialysate. In our motivating dataset, Raman spectra are collected from used dialysate samples collected at different time points during hemodialysis sessions of several ESKD patients.

We seek to develop a statistical method to not only test for difference in Raman spectra collected at different time points during hemodialysis treatment, and thus detect progress in-session. Moreover, our method can detect regions along the wavelength of Raman spectra, and help inform physicians of chemical components that are not filtered out by the hemodialysis process.

1.1.2 Degradation analysis of rechargeable batteries through Voltage discharge curves

Chapter 4 of this dissertation explores the application of functional data data analysis to reliability data analysis, specifically degradation studies. This is motivated by a life testing study from the NASA Ames Prognostics Data Repository [54]. In this study, NASA scientists continuously charge and discharge rechargeable Lithium-ion batteries under different experimental conditions, namely temperature, level of discharge current, and level of stopping voltage. Throughout charge and discharge cycles, detailed data are recorded for each battery through sensors and impedance tests. We are interested in voltage discharge curves (VDC) collected throughout the discharge cycles. Figure 1.2 plots the voltage discharge

curves of the first battery in the dataset.

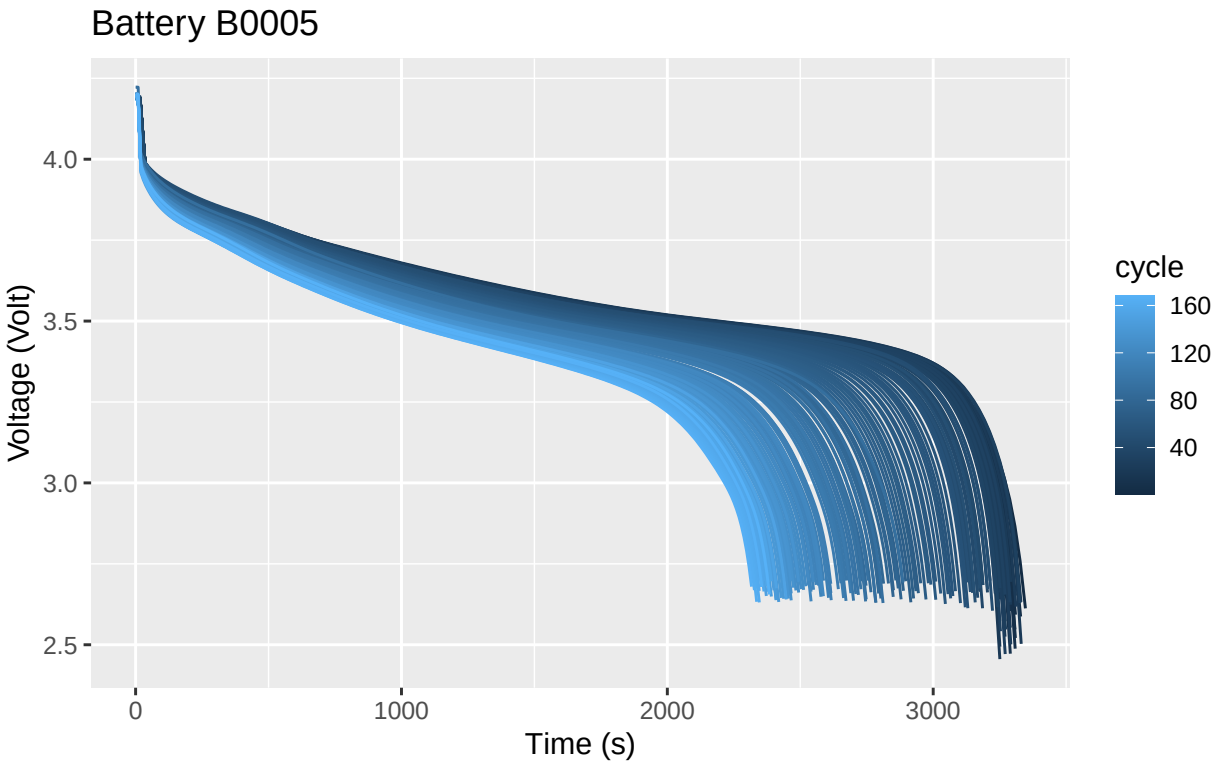


Figure 1.2: Discharge voltage curves throughout discharge cycles of battery B0005

It is not surprising that repeated charge and discharge hasten battery degradation. In Figure 1.2, this degradation can be visualized through the gradient of colors representing the discharge curves over the cycles. While most literature has concentrated on studying battery degradation through the use of a scalar measurement. With the availability of VDCs, we are motivated to take full advantage of the whole discharge curves to learn more about battery performance. It can be seen from Figure 1.2 that each VDC has a shape of a smooth curve, calling for an approach of functional data.

In Chapter 4, we propose a method to study the degradation of rechargeable batteries through VDCs collected longitudinally across discharge cycles. With the incorporation of experimental variables, our method provides insights into battery performance and degradation

under a give condition. More importantly, it provides prediction for battery performance in future discharge cycles in fully functional form, allowing scientists to apply their own degradation measurement and soft failure threshold in monitoring battery health.

1.2 Functional Data Analysis

1.2.1 Functional data

Advancement in technology has enabled data to be collected in higher volume and at finer interval. Due to the continuation of this process, statistical methods that normally operate under the assumption of independence between data points are no longer appropriate for analyzing this type of data. Functional data analysis is developed to deal with this aptly called "functional data", data that are assumed to be generated from an underlying smooth function because of their high dimension nature. In practice, a functional data observation is composed of discrete data values collected over a fine grid. These data points may be contaminated by some amount of noise, but the underlying signal is a smooth curve, and in the case of two or three-dimensional grids, surface or shape.

Since the 1990s, Functional data analysis (FDA), as a field of statistics, has seen a steady increase in interest due to its wide applications. Early applications of FDA include weather and climatology (studying the relationship between daily temperatures and annual precipitation [51]), physiology (longitudinal study of children growth curves [50]), and ergonomics (analysis of body movement [62]). Recently, spectral data are recognised as functional data due to high-dimensional nature. The fact that these data, usually collected over a fine grid of wavelengths and values, reflect some physical and chemical phenomenon makes FDA an ideal method to apply. Several applications of FDA on spectral data have been published

in the field of chemometrics [22], food and nutrition science [58], and medical [60].

The monograph by Ramsay and Silverman [51] includes an overview of functional data analysis together with methods for exploratory and descriptive analysis of functional data. Statistical methods and models that have been established in univariate or multivariate analysis are extended to functional data like functional classification, functional principal component analysis. There is also a class of methods used for models involved both functional and multivariate data. For example, functional multiple regression or ANOVA (functional response and scalar covariates), scalar-on-function regression or functional linear regression (scalar response and functional covariates), function on function regression. General overview of established methods mentioned above as well as recent developments in the field of functional data analysis can be reviewed in [69], [35] and [1].

The following subsections present some background on FDA techniques that are used prevalently throughout this dissertation. Section 1.2.2 covers representing functional observation with basis functions, one of the first steps in the exploration of a functional dataset. Section 1.2.3 covers the essential FDA topic of functional principal component analysis. Sections 1.2.4 and 1.2.5 introduce two popular models within the family of functional linear regression dealing with functional response and scalar response respectively.

1.2.2 Smoothing using basis functions

In reality, functional data almost always come with none or very limited knowledge of the real form of the underlying function $f(t)$. In most cases, a functional datum consists of a set of discrete measures, y_j where j is the index of the sampling points within the data. y_j is the realization of f at a point t_j, contaminated by noises: $y_j = f(t_j) + \epsilon_j$. It is not necessary to recover the true form of f, but it is important to identify a representation of f using the

data on hand so that further analysis can be done. Smoothing is the process of using the observations $y_j, j = 1, 2, \ldots, T$ to find a good representation of f so that the evaluation of $\hat{f}$ at any value t^* can be done and the result is a reasonable estimate of $f(t^*)$.

There is a vast literature on identifying $\hat{f}(t)$ based on the prior assumption on the true form of $f(t)$. For example, if $f(t)$ is assumed to be a linear function, then one can treat the pairs (y_j, t_j) as observations in an ordinary linear regression model and apply the least squares method to obtain the estimate $\hat{f}(t) = \beta_0 + \beta_1 t$. For nonlinear forms of functions, common methods include kernel smoothing, splines, and Gaussian processes. In this dissertation, we concern with representing f through a basis expansion with a roughness penalty.

To perform smoothing with basis functions, we use a set of known functions, $\{\phi_k(t), k = 1, 2, \ldots\}$. This set of known functions constitute a basis system in which all member functions are linearly independent of each other. Other desirable features of basis function are that they should be fast and easy to compute. If desired, basis functions should be sufficiently smooth to accommodate derivatives of various orders. To represent a function f, we form a linear combination of L basis functions from the chosen basis system.

$$f(t) = \sum_{l=1}^{L} c_l \phi_l(t) \tag{1.1}$$

There are many basis systems that have been discovered and utilized for different situations. B-spline basis systems are one of the most popular due to its advantage in efficient computation and compact support property. A set of B-spline basis functions are uniquely identified through a number of, say, M locations (called knots) on the domain and a common order, $d + 1$, for the basis functions. The number of basis functions, L, is then equal to $M + d - 1$. The actual form of B-spline basis functions are quite complicated. Due to its popularity, B-spline functions are built in many statistical packages like R [48], and JMP [47]. The

B-spline function can be evaluated once the order and the number of knots (or specific knot placements) are given. Equation (1.2) displays a recursive format of the i^{th} B-spline function in a B-spline system of order $d+1$ and M knots.

$$\phi_i^d(t) = \frac{t - t_i}{t_{i+d+1} - t_i}\phi_i^{d-1}(t) + \frac{t_{i+d+2} - t}{t_{i+d+2} - t_{i+1}}\phi_{i+1}^{d-1}(t), i = 1, 2, \ldots, L \tag{1.2}$$

with

$$\phi_i^{-1}(t) = \begin{cases} 1 & t_i \leq t < t_{i=1} \\ 0 & \text{otherwise} \end{cases}$$

B-spline functions are not orthogonal of each other. However, B-spline functions are constructed such that the derivatives of the representer at each knot are continuous up to order $(d+1)$. Besides smoothness, another appealing feature of B-splines is that each B-spline basis function is nonzero over at most $d+1$ consecutive intervals or over the intervals formed by $d+3$ adjacent knots. This property is highly advantageous when it comes to representing a function with local sparseness using B-splines.

The representation (1.1) now connects the real data observation $y_j, j = 1, \ldots, T$ and the basis functions through a linear model $y_j = \sum_{l=1}^{L} c_l x_{jl} + \epsilon_i$ where $x_{jl} = \phi_l(t_j)$. This can be rewritten in matrix form to be $\mathbf{y} = \mathbf{X}\mathbf{c} + \epsilon$. To estimate $\mathbf{c}$, we can apply least squares to minimize the mean squared errors $MSE = ||\mathbf{y} - \mathbf{X}\mathbf{c}||_2^2$, leading to the familiar estimate $\hat{\mathbf{c}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$.

A commonly used B-spline family is the cubic splines with $d = 3$. The choice for M is a lot more varied and could have great impact on the final estimates. Too small an M could lead to an overly smoothed function while a large M will produce a wiggly representation

overfitting the raw data. To reduce the dependence on the choice of number of basis functions, a preferred approach is to attach a penalty term to the objective function to penalize the roughness of the function estimate.

$$Q = MSE + \gamma \text{PEN}(f) = ||\mathbf{y} - \mathbf{Xc}||_2^2 + \gamma \text{PEN}(\mathbf{Xc})$$

The penalty, PEN(f) quantifies the roughness of a function f and is often set to be the integrated squared second derivative, $\text{PEN}(f) = \int [D^2 f(t)]^2 dt$. The rougher the representing function $\mathbf{Xc}$ is, the higher penalty PEN($\mathbf{Xc}$) it adds to the objective function Q. Hence, the final estimate is a compromise between a good fit to the observed data and a smooth function. The parameter γ, called the smoothing parameter, controls the tradeoff. When $\gamma = 0$, the roughness of the representing function is disregarded. The larger γ is, the smoother the function estimate will be.

1.2.3 Functional principal component analysis

Functional principal componenet analysis (FPCA) is one of the most popular topics in functional data analysis. The core idea behind FPCA is to represent a group of functional observations as linear combinations of empirical orthogonal basis functions called functional principal components (FPC). While basis function representation in the above section relies on a set of basis function specified a priori, functional principal components are data-driven bases that can be obtained from a group of functional data. To perform FPCA, one starts from a number of functional observations. The goal of FPCA is to extract core function features called functional principal components that are the modes of variations for the functional observations.

FPCA seeks to represent each functional datum $x_i(t)$ as a linear combination of orthogonal FPCs $\xi_k(t)$ weighted by functional scores f_{ki} such that

$$x_i(t) = \mu(t) + \sum_{k=1}^{\infty} f_{ki}\xi_k(t) \tag{1.3}$$

where $\mu(t)$ is the overall mean function. This is guaranteed by the Karhunen-Loève expansion. Note that that for each component i and j, $\xi_i(t)$ and $\xi_j(t)$ are orthogonal, that is $\int \xi_i(t)\xi_j(t)dt = 1$ when $i = j$ and 0, otherwise.

In reality, we observe discrete measures, y_{ij} from a n functional data, $x_i(t_{ij})$, where $i = 1, \ldots, n$ and $j = 1, \ldots, T_i$. That is

$$y_{ij} = x_i(t_{ij}) + \epsilon_{ij} \tag{1.4}$$

The sample mean function for these functional data can be obtained as $\bar{x}(t) = \frac{1}{n}\sum_{i=1}^{n} x_i(t)$ and sample covariance function $\hat{\sigma}(s,t) = \text{cov}(x(s), x(t)) = \frac{1}{n}\sum_{i=1}^{n}\left(x_i(s) - \bar{x}(s)\right)\left(x_i(t) - \bar{x}(t)\right)$. We carry out the FPC decomposition using the sample covariance function $\hat{\sigma}(s,t)$ following the sequential steps. First, we identify $\hat{\xi}_1(t)$ such that $\hat{\xi}_1(t)$ maximizes $\frac{1}{n}\sum_{i=1}^{n}\hat{f}_{1i}^2 = \frac{1}{n}\Sigma_{i=1}^{n}\left(\int \hat{\xi}_1(t)[x_i(t) - \bar{x}(t)]\right)^2 dt$ subject to $\int \hat{\xi}_1^2(t)dt = 1$. Then, we find $\hat{\xi}_2(t)$ to be the maximizer of $\frac{1}{n}\sum_{i=1}^{n}\hat{f}_{i2}^2$ subject to $\int \hat{\xi}_2^2(t)dt = 1$ and the additional constraint for orthogonality $\int \hat{\xi}_2(t)\hat{\xi}_1(t)dt = 0$. $\hat{\xi}_3(t), \hat{\xi}_4(t), \ldots$ are obtained similarly in sequential order. Note that, throughout the algorithm, principal component scores are defined as $\hat{f}_{ki} = \int \hat{\xi}_k(t)[x_i(t) - \bar{x}(t)]dt$. In practice, we usually obtain up to p number of components that explain a specific percentage of total variation in the data, defined as $\sum_{k=1}^{p}\frac{e_k}{\sum_{k=1}^{\infty} e_k}$ where $e_k = \frac{1}{n}\sum_{i=1}^{n} f_{ki}^2$. The final reconstruction of $x_i(t)$ is

$$x_i(t) \approx \bar{x}(t) + \sum_{k=1}^{p} \hat{f}_{ki}\hat{\xi}_k(t)$$

This FPC decomposition could be used in many stages within the analysis of functional data. In data plotting and exploration, the shapes of principal functions help visualize the biggest mode of variations compared to the mean trend. After FPCs and their corresponding scores pairs are obtained, the FPC scores could be used as response variables in predictive models with the goal to forecast functional response. The literature on FPCA is ever expanding. The fundamentals are covered in [51] as explained above. Since then, new FPCA techniques are introduced to adapt to challenges from new and complex types of functional data. For example, FPCA for sparse data in [72], spatially correlated data in [26], and longitudinal functional data in [42].

1.2.4 Functional analysis of variance

The problem of comparing between the means among groups of functional data has been a popular subject in the past decades. One of the earliest work is [14] where testing of difference among means of groups of functional data could be viewed as a regression problem with a functional response and a categorical predictor. The approach is to rewrite the model in a matrix form with the functional response treated as a multivariate vector and the design matrix comprised of indicator variables representing the group labels. The author then took a multivariate analysis approach in estimating the regression coefficients. The test statistic for testing the significance of a full model (using indicator variables for the groups are meaningful) versus a reduced model (intercept only model is better) was estimated at first through a bootstrap method, and later was given a more theoretical treatment in [62].

Taking a more functional approach, Ramsay and Silverman [51] compared the means of k

groups of functional data through the model:

$$y_{ij}(t) = \mu(t) + \alpha_i(t) + \epsilon_{ij}(t), i = 1, 2, \ldots, k; j = 1, 2, \ldots, n_i \tag{1.5}$$

Here, $\mu(t)$ is the grand mean function and $\alpha_i(t)$ is the effect function of group i. Each group has its own size, n_i, the number of curves belonging to group i. The total number of curves in the data is denoted by $N = \Sigma_{i=1}^{k} n_i$. The zero sum constraint is imposed so that $\Sigma_{i=1}^{k} \alpha_i(t) = 0, \forall t$. The model above can be transformed into a functional regression model with the same design matrix Z as that in the classic one-way ANOVA model. The regression coefficient vector $\boldsymbol{\beta}$ consists of functions. The matrix notation of this regression model is

$$\boldsymbol{y}(t) = Z\boldsymbol{\beta}(t) + \boldsymbol{\epsilon}(t). \tag{1.6}$$

The least squares method is used to estimate $\boldsymbol{\beta}$, resulting in the estimate of the coefficient function, $\hat{\boldsymbol{\beta}}(t) = \left(Z^T Z\right)^{-1} Z^T \boldsymbol{y}(t)$. It is not hard to recognize that the value of $\hat{\beta}(t)$ at a specific point t is the classic least square estimate of the coefficient if treating the data at each t value as a separate one-way ANOVA problem. On the other hand, the estimation of the coefficient function $\beta(t)$ can also use a penalization approach with $\beta(t)$ represented by a system of basis functions. The task is now to estimate the coefficients of these basis functions as the optimizer of an objective function formed by combining the MSE and a roughness measurement with a smoothing parameter.

Inference from this model can also be done through the functional version of the familiar tools for evaluating goodness of fit in a one-way ANOVA model. Specifically, the F-ratio function,

$$F\text{-ratio}(t) = \frac{MSR(t)}{MSE(t)} \tag{1.7}$$

This is the functional version of the F-statistic, computed as the ratio between the mean squared regression error, $MSR(t) = \frac{SSY(t)-SSE(t)}{df(regression)}$ and the measure squared error $MSE(t) = \frac{SSE(t)}{df(error)}$. The test of significant difference between the group means of functional data can be done point-wise by comparing the value of the F-ratio at a fixed t to a significant value.

To avoid the naive pointwise multiple tests, researchers have introduced a variety of treatments toward the F-ratio function in (1.7). Cuevas et al. [9] compressed this function into one statistic by calculating the L_2 norm ratio of the numerator and denominator. An asymptotic parametric bootstrap method was implemented to determine the significant level of this test statistic. Zhang et al. [78] considered the test statistic $\sup_{t\in T}$ F-ratio(t) in a procedure named as the "Fmax" test. The null distribution of this test statistic was provided and could be estimated by bootstrap methods. Zhang and Liang [76] proposed a globalizing pointwise F test by integrating F-ratio(t) over the region of interest. They proved that the asymptotic null distribution for this test statistic is a mixture of chi-square distributions with the weights determined by the eigenvalues of the correlation matrix shared by data from all groups.

The functional ANOVA problem has also received treatments from the framework of multivariate analysis, Bayesian analysis, and functional time series. Fan and Lin [12] proposed to test the difference between mean curves from groups of functional data by first reducing dimensionalities of curves either by orthogonal transforms or functional principal component analysis and then applying the high-dimensional ANOVA techniques. Instead of dimension reduction, Antoniadis et al. [2] applied wavelet representations to the ANOVA model and produced an optimal testing method. The work also provided suggestions for ad-hoc contrast tests. Kaufman and Sainy [21] proposed a Bayesian approach to the functional ANOVA problem via imposing a Gaussian process prior on the "batch" function, the analog of the function for group effect. Nguyen and Gelfand [38] offered a Bayesian treatment of the functional ANOVA problem in a non-parametric setting. Functional time series refer to data

where functional observations are correlated over time. Horváth et al. [19] proposed a test of equality for the means of two groups of functional time series. This approach was then extended to the ANOVA setting with more than two groups in Horváth and Rice [18].

A natural question following the testing of difference between group means in the context of functional data is where on the domain the mean functions differ from each other. This question cannot be answered by papers that compressing the F-ratio(t) into a single index by integration or taking supremum. The question has been barely addressed in the context of functional ANOVA. One exception is Pini and Vantini [45], who proposed a procedure for testing the equality of two group means that can also give significant levels at different intervals on the domain. Their method involves three stages: (i) B-spline basis expansion on the data, (ii) testing mean difference from varying-size sets of consecutive basis coefficients from both groups, (iii) multiple comparison correction. This method was then extended to the functional ANOVA setting in Pini et al. [46].

1.2.5 Functional linear model with scalar response

Functional linear model with scalar response is another large group of functional linear regression models. Another name for this type of model is scalar-on-function regression model. This type of regression occurs when the response is a scalar variable and one of the covariates is functional. This model has seen its use in many applications from using the functional temperature profile to predict annual precipitation in [51] to assessing the predictability of feotal heart rates collected longitudinally when treated as a functional covariate toward the continuous response, infant's psychomotor development, in [52]. Ramsay and Silverman [51]

specify this model with the form

$$y_i = \alpha + \int x_i(t)\beta(s)ds + \epsilon_i \quad , i = 1, 2, \ldots, n \tag{1.8}$$

where $\beta(s)$ is the functional covariate presenting the effect of $x_i(t)$ on y_i at point t and ϵ_i is the noise component with mean 0 and variance σ^2. Parameters can be estimated through an objective function with penalized least squares form.

$$Q_\lambda(\beta) = \sum_{i=1}^{n} \left(y_i - \alpha - \int \beta(t)x_i(t)dt \right)^2 + \lambda \int [L\beta(t)]^2 dt \tag{1.9}$$

The first term in (1.9) is the standard least square terms and the second term regularizes the smoothness of $\beta(t)$ through smoothness measurement $L\beta(t)$. In computation, $\beta(t)$ needs to be represeneted by a system of basis functions, $\beta(t) = \sum_{j=1}^{J} c_j\phi_j(t)$. Then, by substibution, object fucntion (1.9) can be re-written in vector-matrix form. The estimated values of α and basis coeffients $\mathbf{c} = (c_1, \ldots, c_J)^\top$ can be obtained in closed-form. Tuning parameter α can be determined through the use of selection criterion such as GCV, BIC, or through cross-validation.

1.3 Reliability Data Analysis

1.3.1 Reliability and failure time analysis

Reliability data analysis is a popular research area in statistics. Its main goal is to study of the reliability of a system. Reliability analysis has its roots of application in physical science fields as opposed to its counterpart, survival analysis, which is more popular in medical and healthcare. One of the first immediate objectives is the analysis and prediction of time to

failure T of a product or a system. In a reliability dataset, the main response variable is failure time with observed values t_i with $i = 1, \ldots, n$, for n units. When the unit does not fail during the data collection period, t_i is called a censored time and unit i is a censored unit.

These types of observations create some challenges. First, time variable needs to be in a positive domain, $T > 0$. Second, the existence of consored observations enforces some complexity in the model construction and estimation. Third, reliability analysis also accounts for situation where time observation for unit i can only be made on a certain domain, $[\tau_i^L, \tau_i^R]$ instead of $[0, \infty]$. This is the situation where data is truncated. Another challenge addressed by reliability analysis is models that accomodate accelerating tests so that system is fast-forwarded to failure time while inference is made in the context of regular use. These challenges require more robust methodology that establish reliability data analysis as its own statistical research field.

To deal with positive observations, methodologies in modeling reliability data rely on probability distributions of positive random variables as opposed to normal random variable. Maximum likelihood estimators (MLE) are usually utilized with the likelihood constructed from the probability distribution functions of positive random variables such as the lognormal, Weibull, log-logistics, and Fréchet distributions. Let $f(t; \boldsymbol{\theta})$ and $F(t; \boldsymbol{\theta})$ respectively denote the probability density function (PDF) and cumulative distribution function (CDF) of the chosen probabiliy distribution where $\boldsymbol{\theta}$ collects the corresponding parameters of the distribution. The PDFs and CDFs of some of the mentioned probability distributions can be found in [6].

We now construct the likelihood function for parameters $\boldsymbol{\theta}$ given the observed data. The contribution of a failure time t_i to the likelihood is $f(t_i; \boldsymbol{\theta})$. To deal with censoring, we utilize indicator variable, δ_i where $\delta_i = 1$ indicates t_i is failure time and $\delta_i = 0$ indicate t_i

is of a censoring type, for example, most popular is right censoring where the unit does not fail when data recording stops. Hence, the contribution of a right-censored observation t_i is $1 - F_(t_i; \boldsymbol{\theta})$ since the failure time for unit i is in the interval $(t_i, +\infty)$. When truncation exists, the observation domain is no longer $[0, +\infty)$, the contribution of the observation to the likelihood needs to be adjusted with a division by the cumulative density over the real observed domain, $\int_{\tau_i^L}^{\tau^R} f(t; \boldsymbol{\theta}) = F(\tau_i^R; \boldsymbol{\theta}) - F(\tau_i^L; \boldsymbol{\theta})$. Hence, the likelihood function taking into account of all observations and their censoring and truncating information has the form of (1.10) for applications with right censoring.

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{f(t_i)_i^{\delta} \left(1 - F(t_i)\right)^{(1-\delta_i)}}{F(\tau_i^R) - F(\tau_i)^L} \tag{1.10}$$

Estimation for parameters $\boldsymbol{\theta}$ can be obtained by maximizing (1.10) or its logarithm form. To account for accelerated testing or situation with covariates $\mathbf{x_i}$, parameters $\boldsymbol{\theta}$ can be represented by a function of the covariate $\mathbf{x_i}$ and alternative parameters to provide interpretation and insights. For example, for log-location-scale distributions like lognormal, Weibull, log-logistic, and Fréchet, the location parameter μ can be rewritten as $\mu = \mathbf{x}_i^{\top}\beta$. By substitution to (1.10), the new parameters become $\boldsymbol{\theta} = (\beta^{\top}, \sigma)^{\top}$ with σ denotes scale parameter. Maximum likelihood estimator for $\boldsymbol{\theta}$ can be obtained similarly.

1.3.2 Degradation analysis

Degradation analysis is another important topic in reliability data analysis. While failure time analysis constructs model on data observed after a system or product has failed, degradation models focus on data collected while product or system is in use. Degradation analysis is becoming more popular for many reasons. First is the advancement in measurement technology that allows data about the system or product to be collected during

operation. Second, as systems become more reliable thanks to continuous improvement and advancement, it is more reasonable to monitor a system's health through its continuing degradation. Study in degradation is also motivated by a practical stand point. Certainly, for some system, it is more sustainable and cost-effective to perform maintenance to slow down the degradation than to replace the entire system due to a failure.

Degradation analysis relies on the collection of data throughout the lifetime of a product. The data collected inform a product's performance or its state of health. Example includes the crack size in materials, the tread depth in tires, and the usage time for batteries. These data have a longitudinal nature with repeated measurements over time. For a system or product where degradation can be tracked, its usability is defined as the period where degradation measure is within a certain threshold. Such a threshold is called the "soft failure". A system is deemed unusable when its degradation measure hits the soft failure. Hence, in degradation analysis, the main goal is to predict the time to the event when the degradation measure meets the pre-specified soft failure. It is also important to obtain the distribution of the time to soft failure for uncertainty quantification.

Typically the prediction of time to soft failure can be obtained through the construction of a degradation path model based on collected data. $\mathcal{D}(t)$ is a function of degradation measure over time. $\mathcal{D}_f$ can be used to represent the degradation measure at which soft failure occurs. Let y_{ij} denote the degradation measurement for unit i at time t_{ij}, $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$. In reality, degradation measure collected for unit i is a realization of the individual degradation path $\mathcal{D}_i(t)$ with some measurement error, $y_{ij} = \mathcal{D}_i(t_{ij}) + \epsilon_{ij}$. Given all the degradation measure and additional covariates $\mathbf{x}_i$, degradation measures y_{ij} can be formally modelled as,

$$y_{ij} = \mathcal{D}_i(t_{ij}; \mathbf{x}_i, \alpha, \gamma_i) + \epsilon_{ij} \tag{1.11}$$

Vector α contains all the common fixed parameters across all the units while $\gamma_i = (\gamma_{1i}, \ldots, \gamma_{im})^\top$

collects m random parameters representing the unit-to-unit variation. In model (1.11), we typically assume γ_i to follow a multivariate normal distribution with mean $\mathbf{0}$ and covariance $\mathbf{\Sigma}$. The measurement errors, ϵ_{ij} are assumed to be identically independent variables following a normal distribution with mean 0 and variance σ_ϵ^2.

The goal now is to estimate the parameters $\boldsymbol{\theta} = \{\boldsymbol{\alpha},\ \sigma_\epsilon^2,\ \mathbf{\Sigma}\}$. Once a functional form of $\mathcal{D}(t)$ is specified, we can see that $z_{ij} = \frac{y_{ij} - \mathcal{D}(t_{ij};\boldsymbol{\alpha},\boldsymbol{\gamma}_i)}{\sigma_\epsilon}$ follows a standard normal distribution. Hence, the likelihood function can be constructed as follows,

$$\mathcal{L}(\boldsymbol{\theta}; \text{Data}) = \prod_{i=1}^{n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\prod_{j=1}^{n_i} \frac{1}{\sigma_\epsilon} \phi_{\mathcal{N}}(z_{ij}) \right] \times \phi_{\text{MVN}}(\gamma_i; \mathbf{\Sigma}) d\gamma_{i1} \ldots d\gamma_{im} \tag{1.12}$$

The maximum likelihood estimates for $\boldsymbol{\theta}$ can be obtained by maximizing (1.12). Notice that the model (1.11) is a nonlinear mixed effects model, which can be estimated through standard methods other than the maximum likelihood. Once the estimated parameters are obtained, we can obtain the CDF for time to soft failure by finding $P[\mathcal{D}(t; \hat{\boldsymbol{\theta}}) \geq \mathcal{D}_f]$ and utilizing the inverse function $T = \mathcal{D}^{-1}(t)$. When this probabiliy is not in a closed form, a numerical CDF can be obtained through simulations. A review of current methods in degradation analysis can be found in [34].

1.4 Contribution of this Dissertation

In my dissertation work, I expand methodologies in Functional Data Analysis under three main projects. The first two projects focus on the setting of a functional ANOVA model. The third project explores the modeling of longitudinally collected functional data in a degradation analysis application for the purpose of rechargeable battery health monitoring.

Functional ANOVA is one of the topics that has garnered a lot of research ever since func-

tional data analysis was established as a statistics field. This speaks to the wide application of comparing groups of functional data. However, much attention has been paid to the testing of difference among groups. In an application like Raman spectroscopy, it is important to not only determine whether there is any difference among groups of functional data but also to identify where on the domain such difference exists and estimate the magnitude of the difference. In the first part of this dissertation, we extend the test of equality between two groups of functional data to a functional contrast test, providing a follow-up analysis tool after the significance of a functional ANOVA test. Another contribution is the expansion of the two-sample test for functional time series developed by [19] into the functional contrast test situation in functional ANOVA model so that the serial correlation among functional observations within each group is accounted for. Next, by incorporating the functional version of smoothly clipped absolute deviation penalty proposed in [24] into the penalized smoothing process of the functional contrast, we produce a sparse representation of the contrast, which both identifies the domain regions where contrast is zero but also gives an estimate of the contrast in the nonzero region.

In the third project, we investigate an application of functional data analysis methodology to reliability and degradation study. Motivated by a battery life testing dataset, we make the case for the use of FDA to take full advantage of functional observations which are the voltage discharge curves (VDC) in our case. Rather than applying an off-the-shelf functional model, we propose a procedure that model and predict functional responses defined on different domains. This can be done through a two-step procedure that begins with scaling each VDC onto the unit domain and save the original domain end time. The domain end time for each cycle is interpreted as the end of discharge time of that cycle in our application. The first step involves modeling the scaled VDCs through FPCA and modeling of the corresponding principal component scores. Second step is the modeling of the domain end time or EOD

time. The two steps together form a functional degradation model (FDM). The proposed FDM incorporates the longitudinal feature of the data via mixed effects models and produces fully functional predictions. Through numerical simulation experiments, we demonstrate that the functional approach has more advantage over an aggregating method in predicting degradation measures. Therefore, it is better equipped for accurate product and system health monitoring, which is the ultimate goal of degradation analysis.

Chapter 2

Contrast Tests for Groups of Functional Data

2.1 Introduction

Testing for difference between groups of functional data has received a lot of attention since the start of functional data analysis as a field of statistics. Suppose k independent groups of functional data are observed: $\{X_{ij}(t) : j = 1, 2, \ldots, n_i\}$, where n_i is the number of observations for the ith group and $X_{ij}(t)$ is an integrable function on a compact interval $\mathcal{T}$. The seminal monograph by Ramsay and Silverman [51] considers the idea of the following functional ANOVA model:

$$X_{ij}(t) = \mu_i(t) + \epsilon_{ij}(t) = \mu(t) + \alpha_i(t) + \epsilon_{ij}(t), i = 1, 2, \ldots, k; j = 1, 2, \ldots, n_i, \tag{2.1}$$

where $\mu_i(t)$ is the group mean function and $\epsilon_{ij}(t)$ are independent random errors with mean zero and covariance functions $\gamma_i(s,t) = \mathrm{Cov}(\epsilon_{i\cdot}(t), \epsilon_{i\cdot}(s))$. Similar to the traditional ANOVA model, the group mean function $\mu_i(t)$ can also be decomposed into the sum of the grand mean function $\mu(t)$ and the effect function $\alpha_i(t)$ of group i. Let n_i be the group size of the ith group, or the number of functional data observations belonging to this group. The total number of functional data observations is denoted by $n = \Sigma_{i=1}^{k} n_i$. The zero sum constraint is

often imposed so that $\Sigma_{i=1}^{k}\alpha_i(t) = 0, \forall t$. Inference from this model can be done through the functional version of familiar tools for a traditional one-way ANOVA model. One important example is the F-ratio function,

$$F\text{-ratio}(t) = \frac{MSR(t)}{MSE(t)}, \tag{2.2}$$

which carries the classic notion of signal to noise, with $MSR(t)$ measuring the strength of the signal and the mean squared error $MSE(t)$ representing the noise. In Ramsay and Silverman [51], the F test is performed point-wise through evaluating the F-ratio at a fixed t and comparing it to the critical value from the corresponding F-distribution. This approach is obviously naïve since it ignores the possible dependence among data points and the multiple testings issue. Therefore, several approaches improving from the F-ratio function have been proposed respectively by Cuevas et al. [9], Zhang et al. [78] and Zhang and Liang [76] to determine the existence of difference among groups of functional data.

When the test is rejected and significant difference among the functional data groups is declared, it is often of practical interest to perform follow-up analysis to determine the pairwise difference among groups of functional data. In some cases, the separations into groups for functional data also carry practical meaning, for example, new treatments versus the standard or placebo treatments. Practitioners may also want to compare the means of functional data from some groups versus those from other groups. This situation calls for a contrast test under the functional ANOVA model. Section 13.3.3 in Ramsay and Silverman [51] introduces a point-wise approach to the contrast test where the test is performed through a permutation procedure. The book by Zhang [75] also includes details on post-hoc contrast testing corresponding to the globalizing point-wise F test [76] and L^2-norm test [78]. A common drawback in these existing contrast tests is that they all require the independence assumption for the observations within a group.

In this paper, we propose a new contrast test that extends the two sample test on functional time series proposed in Horváth et al. [19]. In particular, a χ^2-distributed test statistic is derived based on a weak convergence result on the error process contrast and the Karhunen-Loève expansion of the covariance process. Similar to the two-sample test in Horváth et al. [19], our contrast test allows dependence between functional observations in the same group. In addition to evaluating the performance of the new contrast test, we also conduct an extensive simulation study to compare the empirical performance of all these different contrast tests under the functional ANOVA model, aiming to provide a practical guide on the choice among these tests under different scenarios. We consider functional data generated from two types of mean functions: common smooth functions and rougher functions resembling spectra. For each type of functions, we consider two types of within-group dependence, Brownian bridge and functional AR(1), and three settings of sample sizes. In addition, both the scenario where all groups have a common covariance and the scenario where groups have difference covariances are studied. All the test procedures are compared in terms of their size control and power.

The rest of the paper is organized as followed. We first give details on the existing contrast tests and the proposed test. Description and results of the simulation studies for comparing these tests are presented next. Then we give details on application of these tests on two real dataset and end with some conclusions.

2.2 Functional Contrast Tests

Under model (2.1), a functional linear contrast is a linear combination of the mean functions: $\ell(t) \equiv \sum_{i=1}^{k} a_i \mu_i(t)$, where the coefficient vector $\boldsymbol{a} = (a_1, a_2, \ldots, a_k)^\top$ satisfies the zero sum constraint $\sum_{i=1}^{k} a_i = 0$. Similar to the linear contrasts in the traditional ANOVA model,

functional linear contrasts are used to identify which group means are different from each other once the ANOVA test results in a rejection. The test of interest for the functional linear contrast $\ell(t)$ is

$$H_0 : \ell(t) = 0 \text{ v.s. } H_1 : \ell(t) \neq 0, \tag{2.3}$$

where the equality is in the L^2 sense such that $\ell(t) = 0$ means $\int_{\mathcal{T}} \ell^2(t)dt = 0$.

When q contrasts are tested simultaneously, their coefficient vectors $\boldsymbol{a}_m, m = 1, \ldots, q$, can be stacked up to form a $q \times k$ coefficient matrix $\boldsymbol{C}$, which is generally assumed to have a full rank. Then the hypotheses become

$$H_0 : \boldsymbol{C}\boldsymbol{\mu}(t) = \boldsymbol{c}(t) \text{ v.s. } H_1 : \boldsymbol{C}\boldsymbol{\mu}(t) \neq \boldsymbol{c}(t), \tag{2.4}$$

Let $\bar{X}_{i\cdot}(t) = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}(t)$ be the sample mean function for the ith group.

2.2.1 Existing tests

Each of the following tests, collected in [75], except for the point-wise tests, has two versions: one built under the assumption of a common covariance function for all the groups, that is, $\gamma_i(s,t) = \gamma(s,t), i = 1, 2, \ldots, k$, and the other without this assumption. In the following we describe for each test its common covariance function version first and then the version without the common covariance assumption.

Point-wise tests

The test statistic is

$$F_n(t) = \frac{SSH(t)/q}{SSE(t)/(n-k)} \tag{2.5}$$

where $SSE_n(t) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij}(t) - \hat{\mu}_i(t))^2$ and $SSH_n(t) = [\boldsymbol{C}\hat{\boldsymbol{\mu}}(t) - \boldsymbol{c}(t)]^\top (\boldsymbol{CDC}^\top)^{-1/2}$ $[\boldsymbol{C}\hat{\boldsymbol{\mu}}(t) - \boldsymbol{c}(t)]$. Here, $\boldsymbol{D} = \text{diag}(1/n_1, \ldots, 1/n_k)$, and $\hat{\boldsymbol{\mu}}(t) = (\hat{\mu}_1(t), \ldots, \hat{\mu}_k(t))^\top$ are the estimators of the group mean functions. An example of $\hat{\mu}_i(t)$ is the unbiased estimator by the sample group mean $\bar{X}_{i\cdot}(t)$.

The test is carried out point-wise through the null distribution $F_n(t) \sim F_{q,n-k}$ for a fixed t. Similar to the point-wise functional ANOVA test, this approach does not give a global conclusion for (2.4). A permutation-based version of the test is described in Section 13.3.3 of Ramsay and Silverman [51]. However, our numerical experiments with it, not shown here, suggest that it is a very time consuming approach and often results in low test sizes.

L^2-norm based test

The L^2-norm based test statistic, inspired by the ANOVA test statistics in [14] and [77], is

$$T_n = \int_{\mathcal{T}} SSH_n(t)dt \tag{2.6}$$

Let $\beta = \frac{\text{tr}(\gamma^{\otimes 2})}{\text{tr}(\gamma)}$ and $\kappa = \frac{\text{tr}^2(\gamma)}{\text{tr}(\gamma^{\otimes 2})}$, where $\text{tr}(\cdot)$ is the trace function defined as $\text{tr}(f) = \int_{\mathcal{T}} f(t,t)dt$ and $\otimes$ is the Kronecker product of bivariate functions, $(f_1 \otimes f_2)(s,t) = \int_{\mathcal{T}} f_1(s,u)f_2(u,t)du$. Under the null hypothesis in (2.4), the test statistic T_n can be approximated by $\hat{\beta}\chi^2_{q\hat{\kappa}}$. Two versions of estimators for β and κ are given. Both use the estimator of $\gamma(s,t)$,

$$\hat{\gamma}(s,t) = \frac{1}{n-k} \sum_{i=1}^{k} \sum_{j=1}^{n_i} [X_{ij}(s) - \bar{X}_i(s)][X_{ij}(t) - \bar{X}_i(t)]$$

In the first version, regarded as the naive version, $\hat{\beta} = \frac{tr(\hat{\gamma}^{\otimes 2})}{tr(\hat{\gamma})}$, $\hat{d} = (l-1)\hat{\kappa}$, and $\hat{\kappa} = \frac{\text{tr}^2(\hat{\gamma})}{tr(\hat{\gamma}^{\otimes 2})}$. A second version of estimates for β and κ, under the Gaussian assumption for the functional

observations, are $\hat{\beta} = \frac{\widehat{tr(\gamma^{\otimes 2})}}{\widehat{tr(\gamma)}}$ and $\hat{\kappa} = \frac{\widehat{\text{tr}(\gamma)}^2}{\widehat{\text{tr}(\gamma^{\otimes 2})}}$ where

$$\widehat{\text{tr}(\gamma^{\otimes 2})} = \frac{(n-k)^2}{(n-k-1)(n-k+2)}\left(\text{tr}(\hat{\gamma}^{\otimes 2}) - \frac{\text{tr}^2(\hat{\gamma})}{n-k}\right) \tag{2.7}$$

and

$$\widehat{\text{tr}^2(\gamma)} = \frac{(n-k)(n-k+1)}{(n-k-1)(n-k+2)}\left(\text{tr}^2(\hat{\gamma}) - \frac{2\text{tr}(\hat{\gamma}^{\otimes 2})}{n-k+1}\right) \tag{2.8}$$

When used in in test of equality among group means in functional ANOVA model, these estimates are shown to have reduced biases when compared with their naive version [15, 75].

When assumption of a common variance function among groups of functional data cannot be guaranteed, we need to estimate the covariance function $\gamma_i(s,t)$ of each group i separately. The null distribution of the test statistic (2.6) becomes $T_n \sim \hat{\beta}_1 \chi^2_{\hat{d}_1}$ where $\hat{\beta}_1 = \hat{B}_1/\hat{A}_1$ and $\hat{d}_1 = \frac{\hat{A}_1^2}{\hat{B}_1}$. In the naive estimate, $\hat{A}_1 = \sum_{i=1}^k a_{n,ii}\text{tr}(\hat{\gamma}_i)$, and $\hat{B}_1 = \sum_{i=1}^k \sum_{j=1}^k a_{n,ij}^2 \text{tr}(\hat{\gamma}_i \otimes \hat{\gamma}_j)$ with $a_{n,ij}$ being the (i,j) element in the $k \times k$ matrix $\boldsymbol{A}_n = \boldsymbol{D}_n^{1/2}\boldsymbol{C}^\top(\boldsymbol{C}\boldsymbol{D_n}\boldsymbol{C}^\top)^{-1}\boldsymbol{C}\boldsymbol{D}_n^{1/2}$.

For the bias-reduced version under the Gaussian assumption, one can use

$$\hat{A}_1^2 = \sum_{i=1}^k a_{n,ii}^2 \text{tr}^2(\hat{\gamma}_i) + \sum_{i \neq j} a_{n,ii}a_{n,jj}\text{tr}(\hat{\gamma}_i)\text{tr}(\hat{\gamma}_j) \tag{2.9}$$

and

$$\hat{B}_1 = \sum_{i=1}^k a_{n,ii}^2 \widehat{\text{tr}(\gamma_i^{\otimes 2})} + \sum_{i \neq j} a_{n,ij}^2 \text{tr}(\hat{\gamma}_i \otimes \hat{\gamma}_j) \tag{2.10}$$

where the unbiased estimators for $\text{tr}^2(\gamma_i)$ and $\text{tr}(\gamma_i^{\otimes 2})$ can be obtained using (2.7) and (2.8) respectively with n replaced by the corresponding n_i and k set to 1.

F-type test

This test transforms (2.5) in a way similar to how [62] and [74] transform (2.2), and results in the test statistic

$$F_n = \frac{\int_{\mathcal{T}} SSH_n(t)dt/q}{\int_{\mathcal{T}} SSE_n(t)dt/(n-k)} \tag{2.11}$$

Under the null hypothesis in (2.4), one has $F_n \sim F_{q\hat{\kappa},(n-k)\hat{\kappa}}$ approximately. Here $\hat{\kappa}$ can be computed from the naive method or the biased-reduced method described above.

The F-type test statistic with different group-wise covariance functions is $F_n = \frac{T_n}{S_n}$, where $S_n = \sum_{i=1}^{k} a_{n,ii}\mathrm{tr}(\hat{\gamma}_i)$. Its asymptotic null distribution is $F_{\hat{d}_1,\hat{d}_2}$ with $\hat{d}_2 = \frac{\hat{A}_1^2}{\hat{B}_2}$ and $\hat{d}_1$ defined as in the asymptotic distribution of the L^2-norm based test statistic T_n. The naive and bias-reduced version of $\hat{A}_1^2$ could be obtained as above. The naive and biased-reduced versions of $\hat{B}_2$ are respectively $\sum_{i=1}^{k} \frac{a_{n,ii}^2}{n_i-1}\mathrm{tr}(\hat{\gamma}_i^{\otimes 2})$ and $\sum_{i=1}^{k} \frac{a_{n,ii}^2}{n_i-1}\widehat{\mathrm{tr}(\gamma_i^{\otimes 2})}$.

Bootstrap tests

When the sample sizes for groups are small or when the data deviate much from Gaussian distributions, Zhang [75] recommends a bootstrap approach instead of using the asymptotic distributions. In this approach, bootstrap samples are randomly drawn from the data and either the L^2-norm based test statistic (2.6) or the F-based test statistic (2.11) can be computed accordingly for each bootstrap sample. The statistic obtained from real observations is then compared with these statistic values from bootstrap samples to obtain an empirical p-value for the hypothesis test.

2.2.2 A new functional linear contrast test

Model, test Statistic, and its distribution

For simplicity, we only illustrate the case of testing one functional linear contrast $\ell(t)$ with the notion that the scenario of multiple contrasts can be easily generalized. An unbiased estimator of the functional contrast $\ell(t)$ is $\hat{\ell}(t) = \sum_{i=1}^{k} a_i \bar{X}_{i\cdot}(t)$. Notice that

$$\sum_{i=1}^{k} a_i \bar{X}_{i\cdot}(t) = \sum_{i=1}^{k} a_i \mu_i(t) + \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{a_i}{n_i} \epsilon_{ij}(t) \tag{2.12}$$

Its L^2-norm is $\int \left(\sum_{i=1}^{k} a_i \bar{X}_{i\cdot}(t) \right)^2 dt = \int \left(\sum_{i=1}^{k} a_i \mu_i(t) + \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{a_i}{n_i} \epsilon_{ij}(t) \right)^2 dt$, which reduces to $\int \left(\sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{a_i}{n_i} \epsilon_{ij}(t) \right)^2 dt$ when H_0 is true. According to Theorem 1 in Horváth et al. [19], we have $\frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} \epsilon_{ij}(t) \xrightarrow{d} Z_i$, where Z_i is the Gaussian process in L^2 with mean 0 and covariance function $c_i(t, s)$.

Consequently, the between-group independence of the processes gives

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{a_i}{n_i} \epsilon_{ij}(t) \xrightarrow{d} GP\left(0, \sum_{i=1}^{k} \frac{a_i^2}{n_i} c_i(t, s)\right) \tag{2.13}$$

Taking the L^2-norms of the processes on both sides of (2.13) yields the weak convergence

$$\int \left(\sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{a_i}{n_i} \epsilon_{ij}(t) \right)^2 dt \xrightarrow{d} \int \Gamma^2(t) dt \tag{2.14}$$

where $\Gamma(t) = GP\left(0, \sum_{i=1}^{k} \frac{a_i^2}{n_i} c_i(t, s)\right)$.

We now derive our test statistics based on this convergence result.

Let $d(s, t) = \sum_{i=1}^{k} \frac{a_i^2}{n_i} c_i(t, s)$ be the covariance function of the Gaussian process in (2.13).

Note that $d(s,t)$ defines a covariance operator D on the space $L^2(\mathcal{T})$ such that $D(f) = \int f(s)d(t,s)ds$ for $f \in L^2(\mathcal{T})$.

Then by the Karhunen-Loève expansion, $\int \Gamma^2(t)dt = \sum_{l=1}^{\infty} \lambda_l N_l^2$, where $\lambda_1 \geq \lambda_2 \geq \ldots$ are non-negative eigenvalues of D and $\{N_l, 1 \leq l \leq \infty\}$ are independent standard normal random variables. Let φ_l be the eigenfunction corresponding to λ_l. Then $D(\varphi_l) = \int \varphi_l(s)d(t,s)ds = \lambda_l\varphi_l(t)$. The estimator for $d(s,t)$ is $\hat{d}(s,t) = \sum_{i=1}^{k} \frac{a_i^2}{n_i}\hat{c}_i(t,s)$ where $\hat{c}_i(t,s) = \hat{\gamma}_{i0}(t,s) + \sum_{j=1}^{n_i-1} K(\frac{j}{h})\{\hat{\gamma}_{ij}(t,s) + \hat{\gamma}_{ij}(s,t)\}$. $\hat{\gamma}_{ij}(t,s)$ is the maximum likelihood estimator of the correlation function $\gamma_{ij}(t,s)$, $\hat{\gamma}_{ij}(t,s) = \frac{1}{n_i}\sum_{k=j+1}^{n_i} \left(X_{ik}(t) - \bar{X}_{i\cdot}(t)\right)\left(X_{i,k-j}(s) - \bar{X}_{i\cdot}(s)\right)$ with $0 \leq j \leq n_i - 1$. Here $K(.)$ sets the weight of the contribution from covariance of subsequent functional observations of the same group. Input for $K(.)$ is an index distance normalized by smoothing bandwidth $h = h(n_i)$ satisfying $h(n_i) \to \infty$, $h(n_i)/n_i \to 0$ as $n_i \to \infty$. Additional conditions for $K(.)$ are continuity, boundedness, $K(0) = 1$ and $K(u) = 0$ if $|u| > c$, for some $c > 0$.

Now let $\hat{D}$ be the covariance operator induced by $\hat{d}(s,t)$ such that $\hat{D}(f) = \int f(s)\hat{d}(t,s)ds$. Then the eigenvalues and eigenfunctions of $\hat{D}$ can be used to estimate the eigenvalues and eigenfunctions of D. That is, the estimators $\hat{\lambda}_l$ and $\hat{\varphi}_l$ satisfy $\hat{D}(\hat{\varphi}_l) = \int \hat{\varphi}_l(s)\hat{d}(t,s)ds = \hat{\lambda}_l\hat{\varphi}_l(t), l = 1, 2, \ldots$.

Therefore, we can approximate the limiting random variable $\int \Gamma^2(t)dt$ in (2.14) by $\sum_{l=1}^{p} \hat{\lambda}_l N_l^2$ with p chosen to be large enough so that $\sum_{l=1}^{p} \hat{\lambda}_l N_l^2$ accounts for a large portion of $\sum_{l=1}^{\infty} \hat{\lambda}_l N_l^2$.

On the other hand, recall that from (2.12) we have $\sum_{i=1}^{k} a_i\bar{X}_{i\cdot}(t) = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \frac{a_i}{n_i}\epsilon_{ij}(t)$ under H_0, whose asymptotic null distribution in (2.13) is $GP(0, d(t,s))$. Consequently, we have $E\left[\left\{\sum_{i=1}^{k} a_i\bar{X}_{i\cdot}(t)\right\}^2\right] \to d(t,s)$ as $N \to \infty$.

Now consider the projection of $\sum_{i=1}^{k} a_i\bar{X}_{i\cdot}(t)$ onto the space spanned by the orthonormal eigenfunctions $\varphi_1, \varphi_2, \ldots, \varphi_p$. Let $b_l = \langle\sum_{i=1}^{k} a_i\bar{X}_{i\cdot}(t), \varphi_l\rangle$, which can be estimated by $\hat{b}_l =$

$\langle \sum_{i=1}^{k} a_i \bar{X}_{i\cdot}(t), \hat{\varphi}_l \rangle$.

Let $\mathbf{b} = (b_1, b_2, \ldots, b_p)^\top$ and $Q = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$. Then $\mathbf{b} \xrightarrow{d} N_p(\mathbf{0}, Q)$. Define the test statistic $U^{(1)} = \hat{\mathbf{b}}^\top \hat{\mathbf{b}} = \sum_{i=1}^{p} \hat{b}_i^2$. Based on (2.14), we have $U^{(1)} \xrightarrow{d} \sum_{i=1}^{p} \lambda_i N_i^2$ under H_0, where $N_i, 1 \leq i \leq p$, are independent standard normal random variables. A normalized version of test statistic is $U^{(2)} = \sum_{i=1}^{p} \frac{\hat{b}_i^2}{\hat{\lambda}_i}$ whose asymptotic null distribution is simply the chi-squared distribution with p degree of freedom, or $\chi^2(p)$.

2.3 Simulation

In simulation studies, we consider two types of contrast function: one of a common smooth function form and the other of a rougher form resembling spectral data that we have in the application. Without loss of generality, three groups are considered and the functional linear contrast of interest has coefficients $\mathbf{a} = \left(1, -\frac{1}{2}, -\frac{1}{2}\right)^\top$, that is, we test on the hypotheses

$$H_0 : \mu_1(t) = \frac{\mu_2(t) + \mu_3(t)}{2} \quad \text{vs.} \quad H_1 : \mu_1(t) \neq \frac{\mu_2(t) + \mu_3(t)}{2} \tag{2.15}$$

For all the scenarios, the new contrast test is compared with the existing ones in terms of the size control and power for a hypothesis test. The nominal levels of all the tests are set to be $\alpha = 0.05$.

We shall use the following abbreviations for all the tests evaluated in the simulations. L2N, L2B, and L2b represent respectively the naive, bias-reduced, and bootstrap versions of the L^2-norm based tests. FN, FB, and Fb are respectively the naive, bias-reduced, and bootstrap versions of the F-type tests. FT2-95 is our proposed test based on functional time series using the standardized test statistic from principal components that explain at least 95% of variation in the data. Similarly, FT2-85 refers to the proposed test where 85% data variation

is used to set the chosen number of principal components.

2.3.1 Simulation settings

Setting 1: True contrast is a common-form smooth function

For this setting, we consider the mean functions $\mu_1(t) = f(t) + g(t)$, $\mu_2(t) = \frac{1}{2}f(t) + g(t)$ and $\mu_3(t) = d_3 f(t) + g(t)$ on the domain $[0, 1]$, where $g(t) = \sin\left(8\pi(t + \frac{3}{40})\right)$, $f(t)$ is a continuous function with three pieces, equal to $2(1-t)\sin(2\pi(t+0.2))$ when $0 \leq t \leq 0.3$, 0 when $0.3 < t < 0.7$, and $2t\sin(2\pi(t-0.2))$ when $0.7 \leq t \leq 1$, and the varying constant d_3 controls the effect size of the contrast difference. In particular, $d_3 = 1.5$ corresponds to H_0 and can be used to assess the size control while all the other values of d_3 can be used to examine the power of a test. In our simulation study, d_3 takes values on a grid ranging from 0.1 to 1.5 with step size 0.2, where a smaller value of d_3 corresponds to a larger deviation of the true contrast from 0. For ease of interpretation, we also calculated a Test Effect calculated as $1 - 0.5(0.5 + d_3)$. With the selected grid values for d_2, the Test Effect ranges from 0 to 0.7. The data are simulated following the model (2.1). We investigate the impact of sample sizes through three different scenarios: small and uniform number of sample curves ($n_1 = n_2 = n_3 = 10$), large and uniform number of sample curves ($n_1 = n_2 = n_3 = 50$), uneven number of sample curves ($n_1 = 10, n_2 = 20, n_3 = 30$). Notice that the sample sizes are quite small in the uneven case. We do not consider the uneven case with large sample sizes because it is very clear from the simulation result that all the tests have high powers in the large sample size case no matter whether the number of curves from each group are equal or not. For the random errors $\epsilon_{ij}(t)$, we consider two settings: i.i.d. Brownian bridge and functional AR(1) (FAR(1)). Note that the second error structure induces dependence among data within the same group. Lastly, we set the variance parameters in the error structure of

each group to be all 1 for the common covariance case, or $1, 1.5$ and 2 for groups 1, 2, and 3 respectively in the different covariance case.

Under each combination of data generation settings (d_3, sample sizes, error type, and error variance), we repeat the process 100 times to examine the power of the test. For a visual display, Figure 2.1 shows the group mean functions of two Test Effect settings and the corresponding randomly generated data with the Brownian Bridge noise structure and two different sets of sample sizes.

Setting 2: True contrast function is a spectrum-like function

Spectra, such as the Raman spectra in our application, are often rougher curves than a common smooth function and thus warrant a separate investigation here. To generate spectrum-like data, we consider the mean functions $\mu_1(t) = 100 \sum_{i=1}^{m} p_i(t)$, $\mu_2(t) = 0.5\mu_1(t)$, and $\mu_3(t) = d_3\mu_1(t)$, on the real data domain $[500, 1523]$, where each p_i is a normal density function centered at a chosen point on the interval $[500, 1523]$ and represents a spectral peak at the location. The height and width of the peak is determined by the standard deviation of the normal distribution. To set the peak locations, we first compute 1024 equally spaced grid points on the interval $[500, 1523]$. We then select $m = 3$ peak locations respectively at the 201^{st}, 501^{st}, and 801^{st} grid points. The standard deviations of the three normal densities are 4, 12, and 20. This is to simulate the varying degrees of peak heights. Similar to setting 1, the constant d_3, varying over the grid from 0.1 to 1.5 with step size 0.2, determines the magnitude of the functional contrast, with $d_3 = 1.5$ corresponding to the null hypothesis. The sample sizes, error type, and error variance settings are similar to those in setting 1. Each data generating setting is also replicated 100 times.

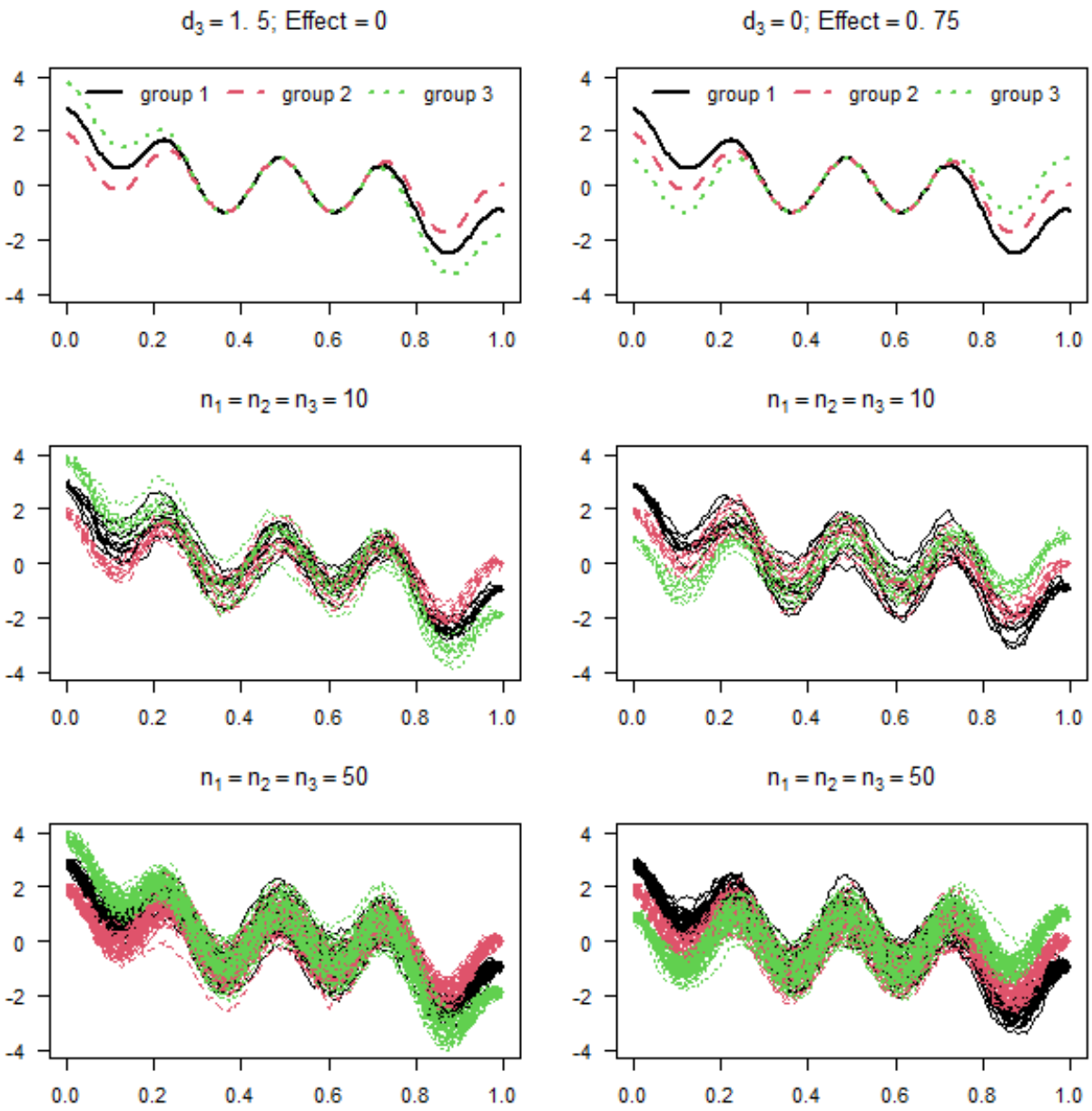


Figure 2.1: Group mean functions and sample functional data under two Test Effect values and two scenarios of sample sizes from Setting 1 of contrast testing simulation.

2.3.2 Result

Tables 2.1 to 2.4 display the empirical rejection percentages of all the tests under the common function simulation settings and Tables 2.5 to 2.8 are for the spectrum-like data settings. In scenarios where the covariance function is set differently for each group, the tests without the common covariance assumption are named after their common covariance versions with an affix "-het" attached to the end of the names. In each table, we present the results for two sample size cases, small ($n_1 = n_2 = n_3 = 10$) and large ($n_1 = n_2 = n_3 = 50$). The case of different sample size ($n_1 = 10, n_2 = 20, n_3 = 30$) gives similar results to the case of ($n_1 = n_2 = n_3 = 10$). Basically, the performance is mostly driven by the size of the observations, not by the difference in group sizes. We refer to the Appendix section A.1 for the display of these results.

Table 2.1: Empirical rejection percentages: the common-form contrast function setting with Brownian Bridge errors and a common group-wise covariance function.

N	N=(10,10,10)								N=(50,50,50)							
Test	L2N	L2B	L2b	FN	FB	Fb	FT2-85	FT2-95	L2N	L2B	L2b	FN	FB	Fb	FT2-85	FT2-95
Effect																
0	7	7	8	6	7	21	11	21	7	7	7	6	7	23	9	10
0.1	10	12	15	10	10	38	38	65	58	60	63	58	58	98	72	100
0.2	56	66	66	45	54	90	59	97	100	100	100	100	100	100	98	100
0.3	99	99	98	99	99	100	85	100	100	100	100	100	100	100	100	100
0.4	100	100	100	100	100	100	90	100	100	100	100	100	100	100	100	100
0.5	100	100	100	100	100	100	95	100	100	100	100	100	100	100	100	100
0.6	100	100	100	100	100	100	95	100	100	100	100	100	100	100	100	100
0.7	100	100	100	100	100	100	97	100	100	100	100	100	100	100	100	100

Table 2.2: Empirical rejection percentages: the common-form contrast function setting with Brownian Bridge errors and different group-wise covariance functions.

N	N=(10,10,10)								N=(50,50,50)							
Test	L2N-het	L2B-het	L2b-het	FN-het	FB-het	Fb-het	FT2-85	FT2-95	L2N-het	L2B-het	L2b-het	FN-het	FB-het	Fb-het	FT2-85	FT2-95
Effect																
0	7	16	8	7	7	3	11	20	4	11	3	4	4	2	7	10
0.1	9	28	9	6	8	6	26	53	25	57	25	24	25	22	54	98
0.2	28	66	28	26	27	19	52	84	100	100	100	100	100	100	94	100
0.3	78	95	79	71	75	58	80	99	100	100	100	100	100	100	99	100
0.4	99	100	98	98	98	90	91	100	100	100	100	100	100	100	100	100
0.5	100	100	100	100	100	100	94	100	100	100	100	100	100	100	100	100
0.6	100	100	100	100	100	100	98	100	100	100	100	100	100	100	100	100
0.7	100	100	100	100	100	100	98	100	100	100	100	100	100	100	100	100

Table 2.3: Empirical rejection percentages: the common-form contrast function setting with FAR(1) errors and a common group-wise covariance function.

N	N=(10,10,10)								N=(50,50,50)							
Test	L2N	L2B	L2b	FN	FB	Fb	FT2-85	FT2-95	L2N	L2B	L2b	FN	FB	Fb	FT2-85	FT2-95
Effect																
0	11	12	12	11	11	29	13	21	12	12	12	12	12	37	7	6
0.1	22	24	22	19	20	49	31	66	69	69	69	67	67	99	52	100
0.2	59	64	61	58	59	92	65	93	100	100	100	100	100	100	97	100
0.3	100	100	97	99	100	100	81	100	100	100	100	100	100	100	100	100
0.4	100	100	100	100	100	100	86	100	100	100	100	100	100	100	100	100
0.5	100	100	100	100	100	100	93	100	100	100	100	100	100	100	100	100
0.6	100	100	100	100	100	100	94	100	100	100	100	100	100	100	100	100
0.7	100	100	100	100	100	100	95	100	100	100	100	100	100	100	100	100

Table 2.4: Empirical rejection percentages: the common-form contrast function setting with FAR(1) errors and different group-wise covariance functions.

N	N=(10,10,10)								N=(50,50,50)							
Test	L2N-het	L2B-het	L2b-het	FN-het	FB-het	Fb-het	FT2-85	FT2-95	L2N-het	L2B-het	L2b-het	FN-het	FB-het	Fb-het	FT2-85	FT2-95
Effect																
0	8	24	11	7	8	6	15	24	12	22	13	11	12	10	8	10
0.1	13	35	14	10	11	7	23	50	28	65	29	27	27	29	33	95
0.2	35	69	34	30	32	22	43	89	100	100	100	100	100	100	79	100
0.3	74	98	75	70	71	63	65	99	100	100	100	100	100	100	99	100
0.4	98	100	98	97	97	93	80	100	100	100	100	100	100	100	100	100
0.5	100	100	100	100	100	100	87	100	100	100	100	100	100	100	100	100
0.6	100	100	100	100	100	100	89	100	100	100	100	100	100	100	100	100
0.7	100	100	100	100	100	100	94	100	100	100	100	100	100	100	100	100

The overall behavior of the tests fits the expected trend of power curves where power is lower close to the null level and increases as the effect moves further away. The powers from all tests also improve with higher sample size. When comparing between the tests, performances are almost indistinguishable among the L2N, L2B, FB, and FN tests. In Setting 1 with the true contrast being a common-form smooth function, the FT2-95 test is mostly the first test to detect significance as the test effect moves away from H_0, and sustains the superior power performance even in the high sample size case. This test, however, suffers from a low size control in the low sample size case. The improvement in the size control of FT2-95 when sample size becomes large is more profound compared to the other tests. The FT2-95 test shows a clear advantage when there is within-group correlation among the functional data (Tables 2.3 and 2.4). It yields high powers in the case of small sample sizes, and achieves better size control while maintaining the highest power with higher sample sizes.

Table 2.5: Empirical rejection percentages: the spectrum-like contrast function setting with Brownian Bridge errors and a common group-wise covariance function.

N	N=(10,10,10)								N=(50,50,50)							
Test	L2N	L2B	L2b	FN	FB	Fb	FT2-85	FT2-95	L2N	L2B	L2b	FN	FB	Fb	FT2-85	FT2-95
Effect																
0	5	5	6	4	5	13	14	26	1	1	1	1	1	14	1	3
0.1	7	8	8	6	7	29	15	35	48	49	51	48	48	100	6	88
0.2	36	39	39	32	34	85	20	53	100	100	100	100	100	100	34	98
0.3	95	97	96	94	95	100	25	71	100	100	100	100	100	100	63	100
0.4	100	100	100	100	100	100	36	88	100	100	100	100	100	100	82	100
0.5	100	100	100	100	100	100	42	90	100	100	100	100	100	100	95	100
0.6	100	100	100	100	100	100	55	95	100	100	100	100	100	100	98	100
0.7	100	100	100	100	100	100	67	96	100	100	100	100	100	100	100	100

Table 2.6: Empirical rejection percentages: the spectrum-like contrast function setting with Brownian Bridge errors and different group-wise covariance functions.

N	N=(10,10,10)								N=(50,50,50)							
Test	L2N-het	L2B-het	L2b-het	FN-het	FB-het	Fb-het	FT2-85	FT2-95	L2N-het	L2B-het	L2b-het	FN-het	FB-het	Fb-het	FT2-85	FT2-95
Effect																
0	7	16	6	7	7	6	18	23	4	10	4	4	4	4	7	11
0.1	8	21	8	7	8	6	18	28	19	52	20	18	18	18	7	68
0.2	16	51	20	16	16	13	19	43	100	100	100	100	100	100	15	97
0.3	61	95	60	54	56	40	20	60	100	100	100	100	100	100	33	98
0.4	94	100	96	93	93	88	30	72	100	100	100	100	100	100	62	99
0.5	100	100	100	100	100	100	39	83	100	100	100	100	100	100	85	99
0.6	100	100	100	100	100	100	41	87	100	100	100	100	100	100	94	100
0.7	100	100	100	100	100	100	46	88	100	100	100	100	100	100	96	100

We conduct additional simulation study to compare size control between existing tests and FT2-95 when group size increases up to $N = (200, 200, 200)$. The result tables in Appendix A.2 demonstrate size control improves the strongest in FT2-95. By $N = (100, 100, 100)$, this test achieves size control close to the nominal level of 5%. These observations also apply to the simulations with spectrum-like true contrast functions although empirical powers are generally smaller than those in the common function setting (Tables 2.5 to 2.8).

Table 2.7: Empirical rejection percentages: the spectrum-like contrast function setting with FAR(1) errors and a common group-wise covariance function.

N	N=(10,10,10)								N=(50,50,50)							
Test	L2N	L2B	L2b	FN	FB	Fb	FT2-85	FT2-95	L2N	L2B	L2b	FN	FB	Fb	FT2-85	FT2-95
Effect																
0	10	13	10	9	9	29	14	22	12	12	12	12	12	33	10	6
0.1	17	21	19	15	17	46	18	29	47	48	50	45	46	100	10	46
0.2	49	54	52	44	48	88	24	47	100	100	100	100	100	100	26	86
0.3	95	99	99	95	95	100	31	57	100	100	100	100	100	100	38	95
0.4	100	100	100	100	100	100	39	71	100	100	100	100	100	100	58	98
0.5	100	100	100	100	100	100	43	80	100	100	100	100	100	100	78	99
0.6	100	100	100	100	100	100	48	86	100	100	100	100	100	100	87	100
0.7	100	100	100	100	100	100	58	87	100	100	100	100	100	100	93	100

Table 2.8: Empirical rejection percentages: the spectrum-like contrast function setting with FAR(1) errors and different group-wise covariance functions.

N	N=(10,10,10)								N=(50,50,50)							
Test	L2N-het	L2B-het	L2b-het	FN-het	FB-het	Fb-het	FT2-85	FT2-95	L2N-het	L2B-het	L2b-het	FN-het	FB-het	Fb-het	FT2-85	FT2-95
Effect																
0	18	26	17	15	16	15	21	23	12	20	12	11	11	11	7	6
0.1	18	32	18	16	17	16	23	34	29	65	27	27	27	25	9	55
0.2	28	54	32	26	27	21	25	43	100	100	100	100	100	100	16	82
0.3	60	98	66	54	58	43	26	54	100	100	100	100	100	100	28	91
0.4	97	100	98	97	97	89	28	67	100	100	100	100	100	100	45	94
0.5	100	100	100	100	100	100	34	80	100	100	100	100	100	100	59	99
0.6	100	100	100	100	100	100	38	85	100	100	100	100	100	100	68	100
0.7	100	100	100	100	100	100	45	88	100	100	100	100	100	100	80	100

2.4 Real Data Example

2.4.1 Raman spectral dataset

Our real spectral data example comes from a study to improve efficacy monitoring of hemodialysis treatment. The treatment is among the most popular for end stage renal disease. Currently, treatments are carried out in a standard procedure for all patients. Patients undergo hemodialysis two to three times per week. Each session typically lasts four hours. The data presented here are collected from real dialysis sessions conducted at hospital facility in collaboration with a biomedical lab from Virginia Tech.

The hemodialysis process involves continuous dialyzing cycles where blood is drawn from patient body into a dialyzer machine where fresh dialysate liquid cleanses metabolic waste

from the blood. The dialyzer then replaces the used dialysate with a fresh batch and returns the cleaned blood to the patient body while starting a new dialyzing cycle. The chemical composition of the used dialysate solution provides critical information about the treatment efficacy that can be extracted by Raman spectroscopy. In this experiment, samples of used dialysate were collected at 5 different time points during real hemodialysis sessions. The time stamps were at 10 minutes, 60 minutes, 120 minutes, 180 minutes, and 240 minutes. Each liquid sample at a time point were divided into a number of portions from which Raman spectrum were generated. Each spectrum from a liquid sample resides over a grid of wavelength where specific wavelength values indicate a specific chemical compound and its concentration in the sample is identified through the intensity on the y-axis. When Raman spectra collected at each time point is considered as a group of functional data, a functional ANOVA test such as the one in [75] can determine whether there is any significant difference among the groups of Raman spectra. A question of more interest is whether the spectra collected during the treatment are significantly different from those collected at time point 1 when the sample would be closer to fresh dialysate. This can be answered through a functional contrast test. Here we consider the comparison of the mean spectra at time points 2 and 3 with the mean spectra at time point 1. All of these posthoc tests can be conducted by any of the methods studied in 2.2.2.

After preprocessing to remove baseline spectrum from the raw spectra data using Iterative Smoothing-spline with Root Error Adjustment (ISREA) [71], we apply the functional ANOVA test to all the five groups and follow up with contrast test on 3 groups of Raman spectra data for time point 1 (10 minutes time stamp), time point 2 (60 minutes), and time point 3 (120 minutes) using the linear contrast $(1, -0.5, -0.5, 0, 0)$.

$$H_0 : \mu_1(t) = \frac{1}{2}\left(\mu_2(t) + \mu_3(t)\right); H_1 : \mu_1(t) \neq \frac{1}{2}\left(\mu_2(t) + \mu_3(t)\right) \tag{2.16}$$

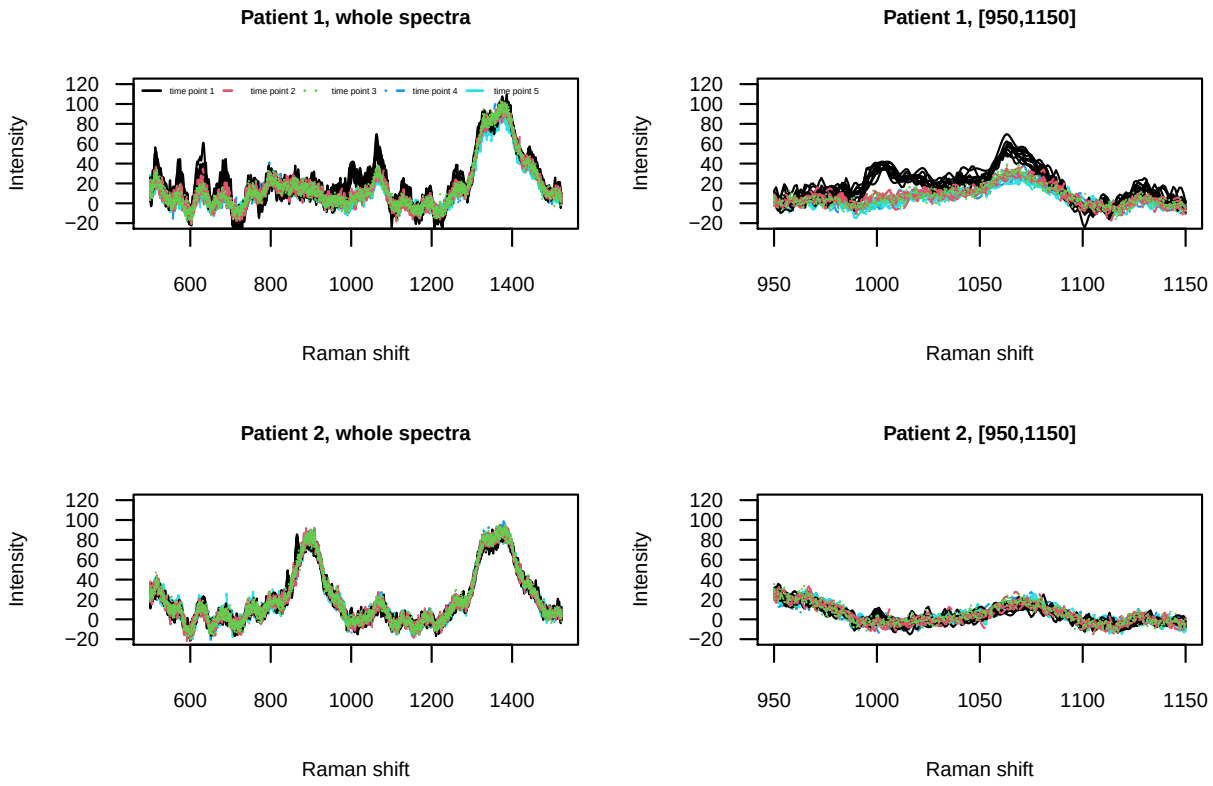


Figure 2.2: Processed Raman spectra from dialysate samples collected at 5 different time points during hemodialysis process of two patients.

Figure 2.2 visualizes the Raman curves among the timepoint groups for two selected patients. The left panels contain the whole spectra while the right panels zoom into the domain from 950 to 1150 Raman shifts where the peak near 1000 marks urea, an important compound in dialysate waste. The results of the contrast test (2.16) for the two patients are presented in Table 2.9. We used the test version where a common variance function is assumed for each spectra group for the L^2-norm based and F-type tests. The data variation for the functional time series test was chosen to be 85%. All the tests rejected the hypothesis for Patient 1 on both the whole spectra and the focused domain $[950, 1150]$. This is not a surprising result given the strong signal displayed in Figure 2.2 by Patient 1. The tests all found that the contrast is not zero for Patient 2 on the whole spectra. However, when focusing on the

domain containing the urea peak, the test based on functional time series did not reject H_0 for this patient.

Table 2.9: Test statistics and p-values (in brackets) for the hypotheses (2.16) on the spectra for three patients.

Patient	L2N	L2B	L2b	FN	FB	Fb	FT2-85
Patient 1, whole spectra	397590.13(0)	397590.13(0)	397590.13(0)	26.86(0)	26.86(0)	26.86(0)	70.31(0)
Patient 1, [950,1150]	234651.72(0)	234651.72(0)	234651.72(0)	77.76(0)	77.76(0)	77.76(0)	187.29(0)
Patient 2, whole spectra	61535.07(0)	61535.07(0)	61535.07(0)	5.27(0)	5.27(0)	5.27(0)	54.25(0)
Patient 2, [950,1150]	10434.55(0)	10434.55(0)	10434.55(0)	4.57(0)	4.57(0)	4.57(0)	16.28(0.13)

2.4.2 Orthosis dataset

The orthosis dataset has been previously studied in [2], [75], and [15]. The data were from a study where subjects wore different types of orthosis devices and performed step-in-place motions from which kinematic forces were recorded. The aim of the study was to compare the effects of different orthopedic conditions on the muscle where perturbation is applied. Each subject performed 10 stepping-in-place replications for each of four conditions: control, with an orthosis device only, and an orthosis device with an additional spring support. There were two types of springs (spring 1 and spring 2). Here we focus on data from a chosen subject of the study. Figure 2.3 displays the raw observations and smooth orthosis curves for the subject under each condition. For this particular subject, the question of interest was whether there is difference between the control condition and the conditions with spring enhancement.

Let $\mu_i(t), i = 1, 2, 3, 4$ denotes the mean force curves under the control, orthosis only, orthosis and spring 1, orthosis and spring 2 conditions, respectively. We're interested in the following contrast test:

$$H_0 : \mu_1(t) = \frac{1}{2}\left(\mu_3(t) + \mu_4(t)\right); H_1 : \mu_1(t) \neq \frac{1}{2}\left(\mu_3(t) + \mu_4(t)\right) \tag{2.17}$$

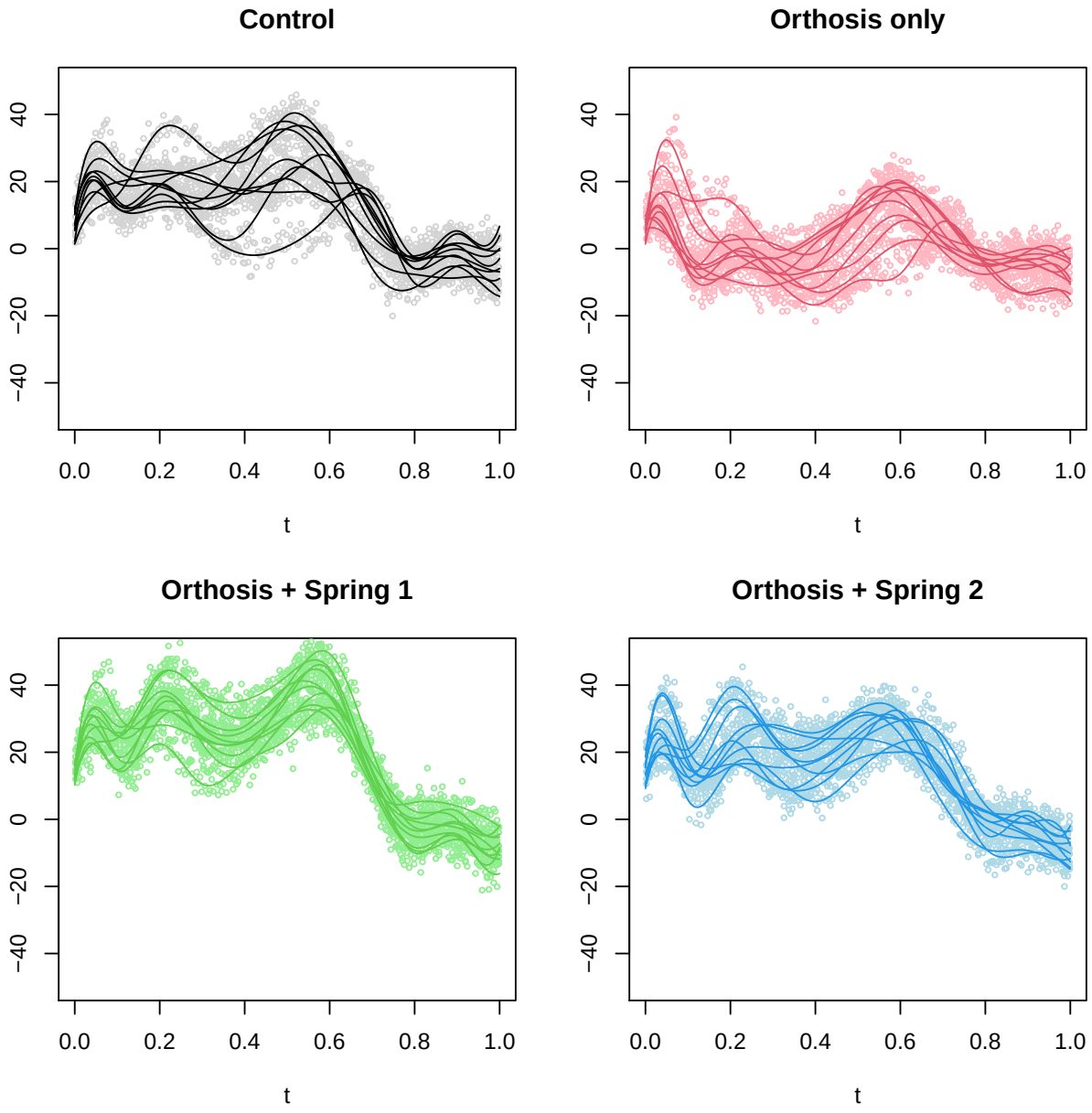


Figure 2.3: Observations and smoothed force curves under four conditions from the Orthosis experiment for Patient 1

From the figure, we can observe the variance of the orthosis curves under each condition are dissimilar. We apply the uncommon variance case for all the tests besides the functional time series test. Results of the contrast tests over two different time regions are presented in table 2.10. All tests reject H_0 over the entire $[0, 1]$ region, and all agree that there is not

enough evidence to reject H_0 over the time frame toward the end of the stepping-in-place action for this patient.

Table 2.10: Test statistics and p-values (in brackets) for the hypotheses (2.17) on two different domains, $t \in [0, 1]$ and $t \in [0.8, 1]$

Contrast	L2N-het	L2B-het	L2b-het	FN-het	FB-het	Fb-het	FT2-95
Control vs springs,[0,1]	64919.27(0)	64919.27(0)	64919.27(0)	4.6(0)	4.6(0)	4.6(0)	39.58(0)
Control vs springs,[0.8,1]	1804.89(0.26)	1804.89(0.13)	1804.89(0.22)	1.35(0.27)	1.35(0.27)	1.35(0.24)	9.14(0.52)

2.5 Conclusion

In this work, we survey existing approaches and introduce a new procedure to perform post-hoc contrast test among groups of functional data under the functional ANOVA framework. The new functional contrast test procedure generalizes the two-sample test procedure in [19] and can account for within-group dependence as well as different covariance functions across groups. We conduct an extensive simulation study to compare these contrast tests when data are generated under two scenarios: smooth common-form functions and functions resembling spiky spectra.

The performance of the existing tests, L2N, L2B, FN, and FB, are similar in all scenarios. These tests give acceptable size control when the null distribution is true. When the number of observations is high and there is correlation between functional observation, we recommend using the test built on functional time series. Even for the independent data case, this test can give stronger or comparable power than the other tests when the truth deviates much from the null hypothesis. This advantage is especially strong for high sample sizes. Hence, we also recommend using this test under these cases.

Chapter 3

Estimating Functional Contrasts with Local Sparsity

3.1 Introduction

Testing procedures for groups of functional data have received a lot of attention as a natural direction for functional data analysis. Suppose k independent groups of functional data are observed: $\{X_{ij}(t) : j = 1, 2, \ldots, n_i\}$, where n_i is the number of observations for the ith group and $X_{ij}(t)$ is an integrable function on a compact interval $\mathcal{T}$. Generalized from the classical ANOVA model, the functional ANOVA model assumes

$$X_{ij}(t) = \mu_i(t) + \epsilon_{ij}(t), i = 1, 2, \ldots, k; j = 1, 2, \ldots, n_i, \tag{3.1}$$

where $\mu_i(t)$ is the mean function of group i, and $\epsilon_{ij}(t)$ are independent random errors with mean zero and covariance functions $\gamma_i(s,t) = \mathrm{Cov}(\epsilon_{i\cdot}(t), \epsilon_{i\cdot}(s))$.

Various inference procedures have been developed based on the functional ANOVA model. Among the first ones is the point-wise F-test introduced in [51], where the test statistic is a straightforward extension of the traditional F-test statistic, with the mean squares calculated at a fixed point on the domain of functional observations. This approach is obviously naïve since it ignores the possible dependence among data points and the multiple testings issue.

Therefore, several improvements have been proposed to determine the existence of difference among groups of functional data since then. Some examples are [9], [78] and [76].

When the functional ANOVA test is rejected and significant difference among the functional data groups is declared, a common follow-up is the functional linear contrast tests that can determine the difference structure among the functional data groups. For example, a point-wise functional contrast test is also introduced in [51] where the test is performed through a permutation procedure. The other monograph Zhang [75] also includes details on post-hoc contrast testing corresponding to the globalizing point-wise F test [76] and L^2-norm test [78]. Chapter 2 extends the two sample tests on functional time series proposed in Horváth et al. [19] to test on functional contrasts, where within-group dependence between functional observations is allowed.

Compared with the contrasts in a classical ANOVA model for scalar responses, a functional contrast is itself a function on the same domain as the functional observations. Upon the declaration of significance for a functional contrast, it is often of practical interest to identify the subregions where the significant difference between relevant groups lies, as shown in our motivating example on hemodialysis monitoring.

Hemodialysis is an important procedure for the treatment of end-stage kidney disease (ESKD). It involves patient blood being continuously pumped into a chamber where a fluid called dialysate is responsible for cleaning chemical waste out of the blood. After the filtration, cleaned blood is transmitted back to the body while the chamber discards the used dialysate and refills with a new batch for the next cycle. Therefore, used dialysate contains critical information in determining the efficacy of the hemodialysis session. Such information can be extracted through Raman spectroscopy. As a well-established method to identify molecular components in a chemical sample, the subjects of study for Raman spectroscopy are Raman spectra. The Raman spectrum of a sample gives a sequence of chemical intensities over a

grid of wavelengths. In our experiment, samples of used dialysate were collected at five time points during a standard hemodialysis session: 10 min, 60 min, 180 min, and 240 min since the start of the session. Each sample was then split into 10 portions. Each portion was scanned by a Raman spectroscope to produce a Raman spectrum. The goal was to monitor the efficacy of dialysis through examining the difference between spectra. When spectra from samples collected at several time points are found to be different from each other by a functional contrast test, it is more interesting to find out where on the spectral domain (i.e., which chemicals) this difference lies. This calls for the identification of the nonzero regions (i.e., local sparsity) of the functional contrast.

As far as we know, local sparsity of functional contrasts has not been studied yet. Pini and Vantini [45] proposed a two-sample function test that not only performs the global testing of difference but also gives significant levels at various intervals along the domain of the functional data. After representing the functional observations by B-spline expansions, they test the mean difference on an interval formed by the supporting intervals of B-spline basis functions through comparing the corresponding consecutive basis coefficients from both groups. This method was later generalized to the functional ANOVA setting in Pini et al. [46]. However, this approach does not really provide a solution to the local sparsity problem. Since the intervals where the mean difference significance levels are calculated are limited by the supporting intervals of B-spline basis functions, these intervals may not give an accurate description of the zero (or nonzero) regions of the contrast function.

To achieve local sparsity goal, we can recognize it as an estimation of the contrast function where the estimate is required to be sparse on regions where the contrast is effectively zero. The notion of "local sparseness" has attracted several works. Wang and Kai [68] proposed the use of group bridge penalty in the estimation of B-splines coefficients in order to produce a sparse estimate for locally sparse function in the context of nonparametric regression

problem. More works on producing locally sparse estimate have been in the estimation of the coefficient function in the functional linear regression problem with scalar response and functional independent variables, $y_i = \beta_0 + \int X_i(t)\beta(t)dt + \epsilon_i, i = 1, 2, \ldots, n$.

The standard approach for the estimation of the coefficient function, $\beta(t)$ is to representing it using basis functions, $\beta(t) = \sum_{l=1}^{L} c_l\phi(t)$. Then the task becomes estimating $\mathbf{c}$ either using least squares method or penalized method to control the smoothness of $\hat{\beta}(t)$. When the true $\beta(t)$ function is exactly zero in some sub-intervals, the estimate from the standard method will still possess some wiggliness and will not entirely be flat over these regions. This is a drawback to the interpretability of the estimated coefficient function since one cannot conclude whether there is an effect of the covariate on the response over these regions. James et al. [20] discretized the domain of $\beta(t)$ into fine points and performed variable selection on selected derivative orders of the bases in mod$B(t)$ using either LASSO or Dantzig selector. They called this methodology "Functional linear regression that's interpretable" (FLiRTI). Zhou et al. [79] point out some drawbacks of FLiRTI in the heavy influence of grid size on the estimate accuracy and the unsatisfying conditions for local sparseness. They proposed an improved method through a two-stage process: initial estimate of sparse regions then further refine. Lin et al. [24], on the other hand, proposed "SLoS" (Smooth and locally Sparse) method by using two types of regularization to achieve both smoothness and sparsity within one optimization procedure. The first penalty is the familiar L2-norm. They extended the "smoothly clipped absolute deviation" (SCAD) penalty by Fan and Li [11] to its functional version and use it to induce local sparsity. The functional SCAD penalty (fSCAD) has since been used in different models such the sparse logistic regression on functional data in [70] and the functional linear regression model with multiple responses in [13]. All of the methods described here attempted to produce a sparse representation for the coefficient function in the functional linear regression model. However, none has been done to directly

detect sparseness in the context of linear contrast from functional ANOVA model.

Our paper is organized as follows: Section 3.2 explains the model and computation methods for contrast testing and sparse contrast detection; Section 3.3 describes and presents results from a simulation studiy on the effectiveness of sparse functional contrast estimation; Section 3.4 gives details of the application of the proposed methods on real dialysis data, followed by conclusion in Section 3.5.

3.2 Method for Estimation Locally Sparse Functional Contrasts

Let $n = \Sigma_{i=1}^{k} n_i$ be the total number of functional observations of all the groups. Recall the functional linear contrast $\ell(t) = \sum_{i=1}^{k} a_i \mu_i(t)$, where the coefficient vector $\boldsymbol{a} = (a_1, a_2, \ldots, a_k)^\top$ satisfies $\sum_{i=1}^{k} a_i = 0$. When the contrast $\ell(t)$ is tested to be significant, it is often of interest to see which part of the domain makes the mean functions differ from each other, that is, on which part of the domain the contrast function $\ell(t)$ is nonzero. Suppose that the functional data are observed at discrete points $t_j, j = 1 \ldots, m$. Then a simple estimate of $\ell(t_j)$ is given by $\tilde{\ell}(t_j) = \sum_{i=1}^{k} a_i \bar{X}_{i\cdot}(t_j)$. Let $y_j = \tilde{\ell}(t_j)$, $\mathbf{y} = (y_1, \ldots, y_m)^\top$, and $\boldsymbol{\ell} = (\ell(t_1), \ldots, \ell(t_m))^\top$. In order to obtain a smooth and locally sparse estimate of $\ell(t)$, we cast the estimation problem into the framework of nonparametric regression and propose to estimate $\ell(t)$ by the minimizer $\hat{\ell}(t)$ of the doubly penalized weighted least squares

$$J(\ell) = \frac{1}{m}(\mathbf{y} - \boldsymbol{\ell})^\top \Sigma^{-1} (\mathbf{y} - \boldsymbol{\ell}) + \gamma \int_{\mathcal{T}} \left\{ \ell^{(s)}(t) \right\}^2 dt + \frac{1}{T} \int_{\mathcal{T}} p_\lambda(|\ell(t)|) dt, \tag{3.2}$$

where the first term is the weighted least squares representing the goodness-of-fit of the estimate, the middle term is a roughness penalty on the estimate, the third term is the

penalty for enforcing local sparsity, and γ and λ are two positive tuning parameters. The covariance matrix Σ is to capture the asymptotic correlations between the raw contrast values y_j. The constant s in the roughness penalty is often chosen to be 2, corresponding to a cubic spline estimator. For the sparsity penalty function $p_\lambda(\cdot)$, we will use the functional SCAD penalty in Lin et al. [24].

A convenient choice of the covariance matrix Σ is the identity matrix, resulting in the so-called working independence procedure, which is known to still yield a consistent mean component (the contrast in our case) estimate despite possibly misspecifying the covariance structure.

A tempting choice of the covariance matrix Σ is an estimate of the asymptotic covariance based on the weak convergence results (2.13) and (2.14). However, our numerical experiments show that this choice actually often deliver much worse contrast estimates than the working independence procedure. Without any additional assumption, a nonparametric estimate of the asymptotic covariance matrix often results in a singular matrix or lacks the accuracy required by the penalized weighted least squares estimation. On the other hand, a good contrast estimate may be achievable if a parametric form of the covariance matrix can be assumed. In Section 3.3, we will study an autoregressive modeling of Σ against the working independence procedure.

Computation

We shall optimize the objective function (3.2) using a B-splines expansion of $\ell(t)$ and a local quadratic approximation to the sparsity penalty term.

Consider representing the contrast function $\ell(t)$ by an expansion of L B-spline basis functions $\phi_1(t), \phi_2(t), \ldots, \phi_L(t)$ of order $d+1$ with $M+1$ equally spaced knots at $0 = t_0 < t_1 < \cdots <$

$t_M = T$ on the interval $\mathcal{T}$. Let $\ell(t_j) \approx \sum_{l=1}^{L} c_l \phi_l(t_j)$ under this representation, where c_l are coefficients. The compact support property of B-splines, that is, each B-spline basis function $\phi_l(t)$ is nonzero only on at most $d+1$ sub-intervals in succession, plays a key role in the local sparsity estimation of $\ell(t)$.

Let Φ be the $m \times L$ matrix with the (j,l)th entry $\phi_l(t_j)$, $\mathbf{c} = (c_1, \ldots, c_L)^\top$, and R be the $L \times L$ matrix with the (j,k)th entry $\int_{\mathcal{T}} D^{(s)}\phi_j(t) D^{(s)}\phi_k(t)dt$. Then the weighted least squares part and the roughness penalty in the objective function (3.2) become respectively $\frac{1}{m}(\mathbf{y} - \Phi\mathbf{c})^\top \Sigma^{-1}(\mathbf{y} - \Phi\mathbf{c})$ and $\gamma \mathbf{c}^\top R \mathbf{c}$.

For the sparsity penalty term, Theorem 1 in Lin et al. [24] shows

$$\frac{1}{T}\int_{\mathcal{T}} p_\lambda(|\ell(t)|)dt = \frac{1}{M}\sum_{j=1}^{M} p_\lambda\left(\frac{\|\ell_{[j]}(t)\|_2}{\sqrt{T/M}}\right), \tag{3.3}$$

where $\|\ell_{[j]}(t)\|_2 = \sqrt{\int_{t_{j-1}}^{t_j} \ell^2(t)dt}$. Note that the right hand side of (3.3) is also quadratic in $\ell(t)$ or $\mathbf{c}$. Therefore, the whole objective function (3.2) is now a quadratic function of $\mathbf{c}$ and can be optimized as follows.

Given the current estimate $\widetilde{\ell}$ of ℓ, we use the local quadratic approximation in Fan and Li [11] to approximate the right hand side of (3.3) as

$$\begin{aligned}
&\sum_{j=1}^{M} p_\lambda\left(\frac{\|\ell_{[j]}(t)\|_2}{\sqrt{T/M}}\right) \approx \\
&\sum_{j=1}^{M}\left[p_\lambda\left(\frac{\|\widetilde{\ell}_{[j]}(t)\|_2}{\sqrt{T/M}}\right) + \frac{1}{2}\cdot\frac{p'_\lambda\left(\frac{\|\widetilde{\ell}_{[j]}(t)\|_2}{\sqrt{T/M}}\right)}{\frac{\|\widetilde{\ell}_{[j]}(t)\|_2}{\sqrt{T/M}}}\left\{\left(\frac{\|\ell_{[j]}(t)\|_2}{\sqrt{T/M}}\right)^2 - \left(\frac{\|\widetilde{\ell}_{[j]}(t)\|_2}{\sqrt{T/M}}\right)^2\right\}\right] = \\
&\frac{1}{2}\sum_{j=0}^{M}\frac{p'_\lambda\left(\frac{\|\widetilde{\ell}_{[j]}(t)\|_2}{\sqrt{T/M}}\right)}{\frac{\|\widetilde{\ell}_{[j]}(t)\|_2}{\sqrt{T/M}}}\frac{\|\ell_{[j]}(t)\|_2^2}{T/M} + G(\widetilde{\ell}),
\end{aligned} \tag{3.4}$$

where

$$G(\widetilde{\ell}) = \sum_{j=1}^{M} p_\lambda \left(\frac{\|\widetilde{\ell}_{[j]}(t)\|_2}{\sqrt{T/M}} \right) - \frac{1}{2} \sum_{j=1}^{M} p'_\lambda \left(\frac{\|\widetilde{\ell}_{[j]}(t)\|_2}{\sqrt{T/M}} \right) \frac{\|\widetilde{\ell}_{[j]}(t)\|_2}{\sqrt{T/M}}$$

On the other hand, $\|\ell_{[j]}\|_2^2 = \int_{t_{j-1}}^{t_j} \ell^2(t)dt \equiv \mathbf{c}^\top \mathbf{W}_j \mathbf{c}$ where $\mathbf{W}_j$ is an L by L matrix with entries $w_{uv} = \int_{t_{j-1}}^{t_j} \phi_u(t)\phi_v(t)dt$ if $j \le u, v \le j + d$ and zero otherwise. Let

$$\widetilde{\mathbf{W}} = \frac{1}{2} \sum_{j=1}^{M} \left(\frac{p'_\lambda \left(\frac{\|\widetilde{\ell}_{[j]}\|_2}{\sqrt{M/T}} \right)}{\|\widetilde{\ell}_{[j]}\|_2 \sqrt{M/T}} \mathbf{W}_j \right) \tag{3.5}$$

Then, the first term on the right hand side of (3.4) can be written as

$$\frac{1}{2} \sum_{j=0}^{M} \frac{p'_\lambda \left(\frac{\|\widetilde{\ell}_{[j]}(t)\|_2}{\sqrt{T/M}} \right)}{\frac{\|\widetilde{\ell}_{[j]}(t)\|_2}{\sqrt{T/M}}} \frac{\|\ell_{[j]}(t)\|_2^2}{T/M} = \mathbf{c}^\top \widetilde{\mathbf{W}} \mathbf{c}$$

Therefore, after dropping the term $G(\widetilde{\ell})$ which does not depend on ℓ or $\mathbf{c}$, the objective function (3.2) can be written fully in the vector-matrix form as

$$Q(\mathbf{c}) = \frac{1}{m} (\mathbf{y} - \Phi \mathbf{c})^\top \Sigma^{-1} (\mathbf{y} - \Phi \mathbf{c}) + \gamma \mathbf{c}^\top R \mathbf{c} + \mathbf{c}^\top \widetilde{\mathbf{W}} \mathbf{c}. \tag{3.6}$$

Since $Q(\mathbf{c})$ is quadratic in $\mathbf{c}$, the updating equation at the the ith iteration can be obtained as

$$\hat{\mathbf{c}}^{(i)} = (\Phi^\top \Sigma^{-1} \Phi + m\gamma R + m\mathbf{W}^{(i-1)})^{-1} \Phi^\top \Sigma^{-1} \mathbf{y} \tag{3.7}$$

Therefore, the complete algorithm for estimating c is

- Step 1: Obtain an initial estimate of $\hat{\mathbf{c}}^{(0)} = (\Phi^\top \Sigma^{-1} \Phi + \gamma R)^{-1} \Phi^\top \Sigma^{-1} \mathbf{y}$. This is the penalized least square estimate of c with no imposition on local sparsity.

- Step 2: Using the estimated coefficients $\hat{\mathbf{c}}^{(i-1)}$ from the previous iteration, compute $\mathbf{W}^{(i)}$ using (3.5). Then, update the estimation using (3.7) to obtain $\hat{\mathbf{c}}^{(i)}$. Set elements of $\hat{c}^{(i)}$ with values smaller than a threshold ϵ to 0.

- Step 3: Repeat Step 2 until the convergence criterion is satisfied.

The convergence criterion in Step 3 is whether the change in $\hat{\mathbf{c}}$ from that of the previous iteration is smaller than a pre-defined threshold. For the tuning parameters γ and λ, we shall study the performance of the Aikaike Information Criteria (AIC), Bayesian Information Criteria (BIC), and Generalized Cross Validation (GCV) defined respectively as

$$\mathrm{BIC}(\lambda,\gamma) = m\log \mathrm{RSS}_{\lambda,\gamma} + (\log m)\mathrm{df}(\gamma,\lambda) \tag{3.8}$$

$$\mathrm{AIC}(\lambda,\gamma) = m\log \mathrm{RSS}_{\lambda,\gamma} + 2\mathrm{df}(\gamma,\lambda) \tag{3.9}$$

$$\mathrm{GCV}(\lambda,\gamma) = \frac{m^{-1}\mathrm{RSS}}{(m^{-1}\mathrm{trace}(\mathbf{I}-\mathbf{S}_{\lambda,\gamma}))^2} \tag{3.10}$$

Here, $\mathrm{RSS}_{\lambda,\gamma} = (\mathbf{y}-\mathbf{\Phi}\hat{\mathbf{c}})^\top\Sigma^{-1}(\mathbf{y}-\mathbf{\Phi}\hat{\mathbf{c}})$ and $\mathrm{df}(\gamma,\lambda)$ is the number of nonzero coefficients in $\hat{\mathbf{c}}$ and $\mathbf{S}_{\lambda,\gamma} = \Phi(\Phi^\top\Sigma^{-1}\Phi + m\gamma R + m\mathbf{W})^{-1}\Phi^\top\Sigma^{-1}$ is the smoothing matrix.

3.3 Simulation Studies

We apply our proposed method for sparse representation of contrast function in two simulations. In each scenario, three groups of functional data are generated based on model (3.1). Two types of functional data are examined. First setting is when data are generated from common smooth functions. In the second setting, we mimic the rough nature of Raman spectra when simulating the data. The purpose of this simulation study is to evaluate the efficacy of the proposed method in smooth estimation and sparse representation of the

contrast function. We compare its performance with the standard roughness penalty smoothing approach. The simulation study also evaluates the performance of different criteria for choosing the tuning parameters γ and λ.

3.3.1 Simulation setting 1: Data from common functions

In this simulation setting, the mean functions for the three functional data groups are respectively $\mu_1(t) = 2f(t) + h(t)$, $\mu_2(t) = \frac{3}{2}f(t) + h(t)$, and $\mu_3(t) = \frac{1}{2}f(t) + h(t)$, where $h(t) = \sin\left(8\pi(t + \frac{3}{40})\right)$ and

$$f(t) = \begin{cases} 2(1-t)\sin\left(2\pi(t+0.2)\right), & 0 \leq t \leq 0.3 \\ 0, & 0.3 < t < 0.7 \\ 2t\sin\left(2\pi(t-0.2)\right), & 0.7 \leq t \leq 1 \end{cases} \tag{3.11}$$

Therefore, the true functional contrast is $\mu_1(t) - \frac{1}{2}\left(\mu_2(t) + \mu_3(t)\right) = f(t)$. The actual functional observations are discretized version of $X_{ij}(t)$ with t from a grid of length 200 spanning $[0, 1]$.

We consider three cases for the distribution of the error term $\epsilon_{ij}(t)$:

1. Gaussian noise, $\epsilon_{ij} \sim MVN(\mathbf{0}, \sigma_\epsilon^2 \mathbb{I}_{T\times T})$
2. AR(1) process, $\epsilon_{ij}(t) = \phi\epsilon_{ij}(t-1) + \delta_t$ with $\delta_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$ and $\phi = 0.8$
3. AR(7) process, $\epsilon_{ij}(t) = \sum_{p=1}^{7} \phi_p \epsilon_{ij}(t-p) + \delta_t$ with $\delta_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$

From examining the spectral data set from our real data application, we find that the mean-subtracted spectral data for each group seem to follow an auto-regressive (AR) process. This is consistent with a recommendation from [51] in that an auto-regressive struc-

ture can be assumed for the residuals. The AR coefficients in case 3 are set to be $\boldsymbol{\phi} = (2.1805, -2.7480, 2.5153, -1.7861, 0.9778, -0.3817, 0.0761)'$, which is an estimate from a randomly selected actual spectral profile.

For each case, the magnitude of σ_ϵ is set to equal a fraction of the standard deviation of the discretized true contrast function. For high noise, this fraction equals $\frac{1}{2}$ in the Gaussian and AR(1) noise case and $\frac{1}{4}$ in the AR(7) case. For low noise, $\frac{1}{4}$ is used for the Gaussian and AR(1) case and $\frac{1}{8}$ is for the AR(7) case. Another aspect to be evaluated is the performance of the method under different observation curves (n_i) for each group. We simulate three scenarios: small sample size ($n_1 = n_2 = n_3 = 10$), large sample size ($n_1 = n_2 = n_3 = 50$), and uneven sample sizes between the groups ($n_1 = 10, n_2 = 15, n_3 = 20$). We generate 100 data sets for each possible combination of noise structures and sample size settings.

We compare our proposed method with the standard approach to smooth contrast function with roughness penalty (abbreviated as **Smooth** in result figures). This is essentially setting the sparsity tuning parameter λ to be 0 and the tuning of the smoothing parameter, γ is done using GCV. Through this simulation study, we also want to evaluate the performance of different selection criteria. Since the ground truth is known, we include the performance when tuning parameters are set to produce a representation with the smallest MSE (abbreviated as **MSE-Best** in result figures). This will not be possible in reality. Thus, fully data-driven selection criteria we consider are **BIC** (3.8), **AIC** (3.9) and GCV. Instead of the GCV expressed in (3.10), we consider an adjusted version,

$$\text{GCV}(\lambda, \gamma) = \left(\frac{m}{m - 1.2\text{df}(\lambda, \gamma)}\right)\left(\frac{\text{SSE}}{m - 1.2\text{df}(\lambda, \gamma)}\right) \tag{3.12}$$

The multiplication by a factor of 1.2 is another layer of discounting as recommended in Gu [16]. Here, $\text{df}(\lambda, \gamma) = \text{trace}[\mathbf{I} - \mathbf{S}_{\gamma,\lambda}]$. However, it is common to use the number of nonzero

coefficients in $\hat{\mathbf{c}}$ in place of $\mathrm{df}(\lambda, \gamma)$. In our simulation we consider both versions of GCV: one using the continuous degree of freedom calculated using trace of $\mathbf{I} - \mathbf{S}_{\gamma,\lambda}$, and one using the discrete number of nonzero coefficients as the degree of freedom. In our reporting, we denote the first version as **GCV-s** and use **GCV-r** to denote the latter version.

The performance of the competitors is evaluated based on the two results it promises to deliver: identification of the zero and nonzero subregions on the domain, and estimation of the magnitude of the functional contrast on nonzero subregions as well as throughout the whole domain. For the first purpose, specific criteria are proportion of nonzero regions that are correctly identified (sensitivity), proportion of null regions that are correctly identified (specificity). Both of these criteria have ranges from 0 to 1. The higher value of each of these two proportions, the better the identification in the corresponding region. To evaluate the estimation quality, we consider the mean spared error on the whole domain, $\mathcal{T}$, $\mathrm{MSE} = \frac{1}{T}\int_{\mathcal{T}} \left(\hat{f}(t) - f(t)\right)^2 dt$; and the integrated squared error defined on the zero subregions $\mathcal{Z}$, $\mathrm{ISE}_0 = \frac{1}{l_0}\int_{\mathcal{Z}} \left(\hat{f}(t) - f(t)\right)^2 dt$; and nonzero subregions $\mathcal{T} \setminus \mathcal{Z}$, $\mathrm{ISE}_1 = \frac{1}{l_1}\int_{\mathcal{T}\setminus\mathcal{Z}} \left(\hat{f}(t) - f(t)\right)^2 dt$. Below we present the results for the estimation quality criteria, MSE, ISE_0, ISE_1. We prefer the readers to the Appendix section B for results on sensitivity and specificity due to their corresponding relationship to ISE_1 and ISE_0 respectively.

Figure 3.1 display the sparse representation of the contrast function under different selection criterion when data are generated from low Gaussian noise case with 10 sample curves for each group. The plots illustrate the advantage of including sparsity regularization in obtaining the representation. No matter what criterion used in selecting the tuning parameters, the estimates for the contrast functions are very good and match up quite well to the true function. While the estimate from using just a roughness penalty is able to replicate the general shape of the true function, it fails to produce a sparse estimates over true zero

regions.

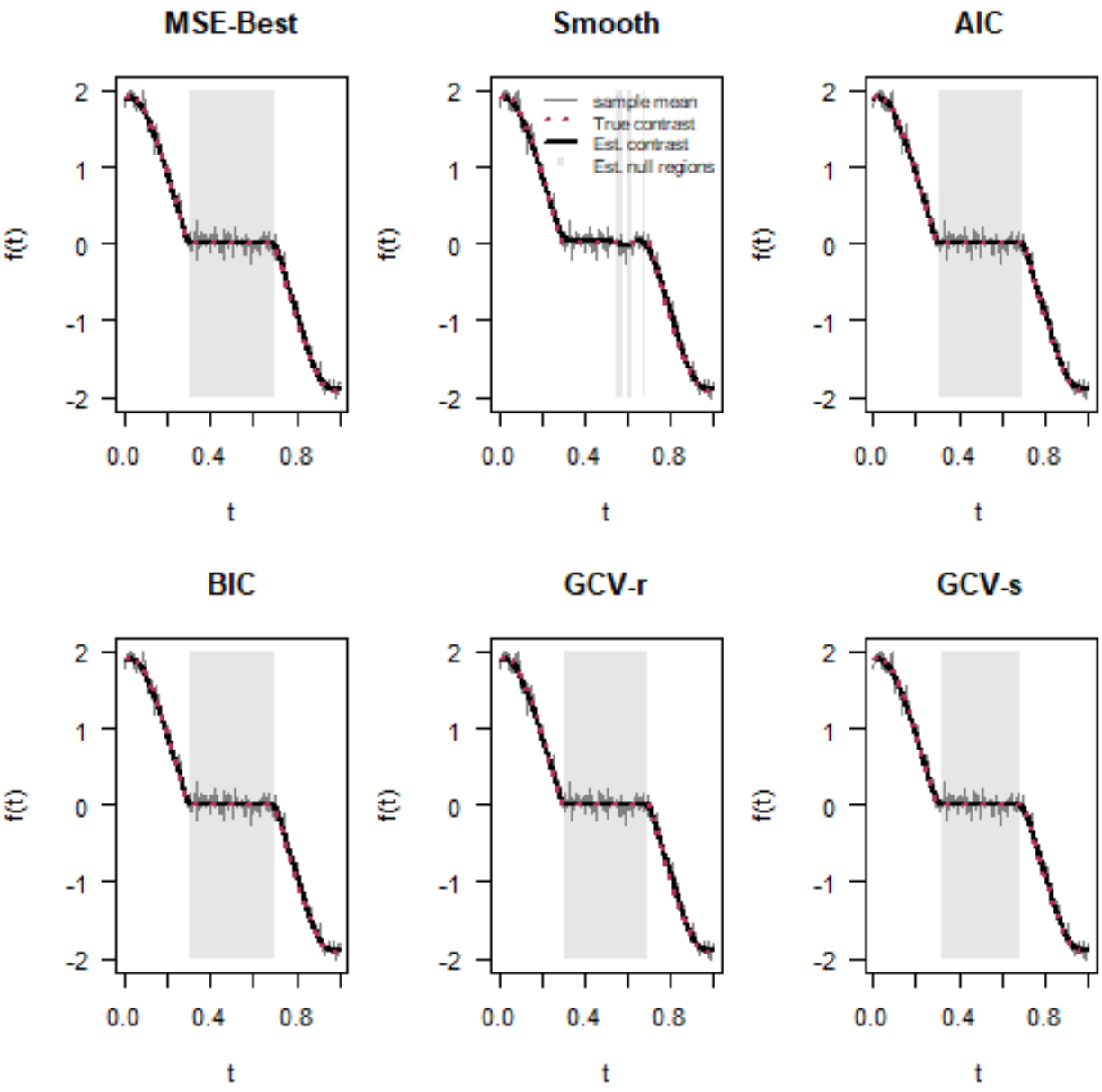


Figure 3.1: Estimates for contrast function of three groups of 10 functional data curves simulated with a common function contaminated with low Gaussian noise case with

Figures 3.2 and 3.4 visualize the performance by criteria in respective performance criterion for data generated by the common functions described in (3.11) and contaminated with Gaussian noise. Some expected results show up from the simulation study. First, the level

of noise play a role in the performance result across the criteria. The error measures (MSE, ISE0, ISE1) are much smaller for the low noise case than the high noise case. Second, the increase in sample size also leads to smaller errors. These two trends are present throughout all data generating mechanisms with different noise structures that we test.

The results from the Gaussian noise case demonstrate that when using a good setting for tuning parameters, the estimate of the contrast function from our sparse representation leads to an improvement over just smoothing with roughness penalty in all of the evaluation criteria. Among the data-driven selection criteria for the tuning parameters, the GCV where the degree of freedom is calculated by smoothing matrix (GCV-s) has performance closest to the best performance by the criteria when the truth is known (MSE-Best). The MSE for fits under this criteria is comparable with that of the fits when tuning parameters chosen under a known function (Figure 3.2). The estimates using tuning parameters chosen by BIC tend to be more accurate in identifying zero regions resulting in very good result with specificity and ISE0 (Figures 3.3 and B.1). However, these fits tend to be on a sparser side, leading to worse performance in terms on sensitivity and ISE1 (Figures 3.4 and B.2). GCV-r performance is similar to that of BIC, but less stable. GCV-s does the best job in balancing the accuracy in both zero and nonzero regions.

When the data is contaminated by non-Gaussian noises, we can apply a weight matrix to improve the fit. This weight matrix can be obtained through AR(p) covariance matrix calculated using estimated AR coefficients and closed-form structure of AR(p) covariance matrix detailed in Brockwell and Davis [3]. For each iteration in the AR(1) and AR(7) case, we include the results from the fits with and without weights across different selection criteria. For the weighted fit, we also compare the performance when the weights are from the true and estimated covariance matrices.

Figures 3.5 to 3.7 display the comparisons in the respective performance criterion for the

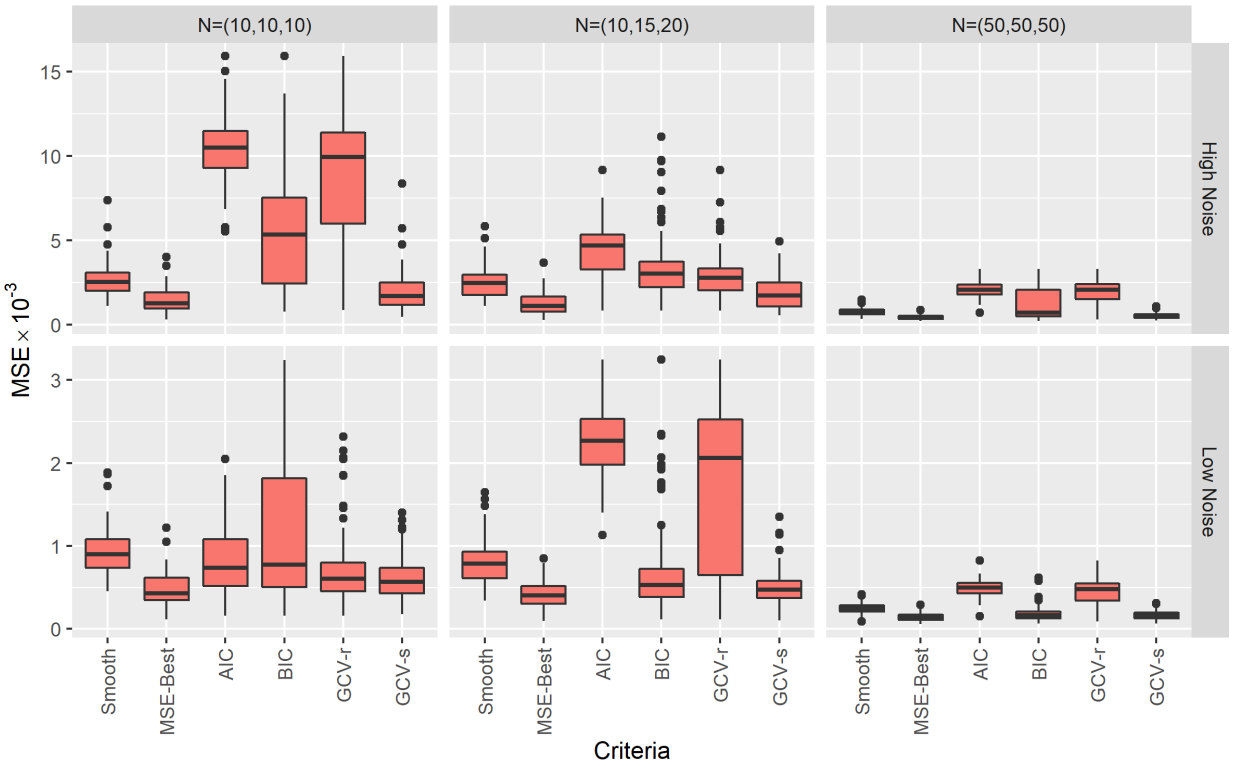


Figure 3.2: Common function case: Mean squared error (MSE) $\times 10^{-3}$ of the smoothed and sparsed contrast function from Gaussian noise case using 7 selection criteria

AR(1) noise case. Figures 3.8 to 3.10 are for AR(7) noise case. Not counting the inaccessible case of MSE-Best (tuning parameters chosen by proximity to the true function), the fits with weight matrix have better results than non-weight version in almost all performance criteria. The results from using estimated covariance matrix are comparable to those using the true matrix.

Figures 3.5 to 3.10 show that GCV-s have the best performance in terms of MSE compared to other data-driven selection criteria. Another common result between AR(1) and AR(7) noise case is that BIC and GCV-r tend to perform well in ISE0 and specificity, but not quite so with ISE1 and sensitivity. This is due to the tendency of BIC and GCV-r to overestimate the length of a sparse region leading to very accurate identification of zero regions, but low performance with non-zreo regions as a trade off. This observation can be explained further

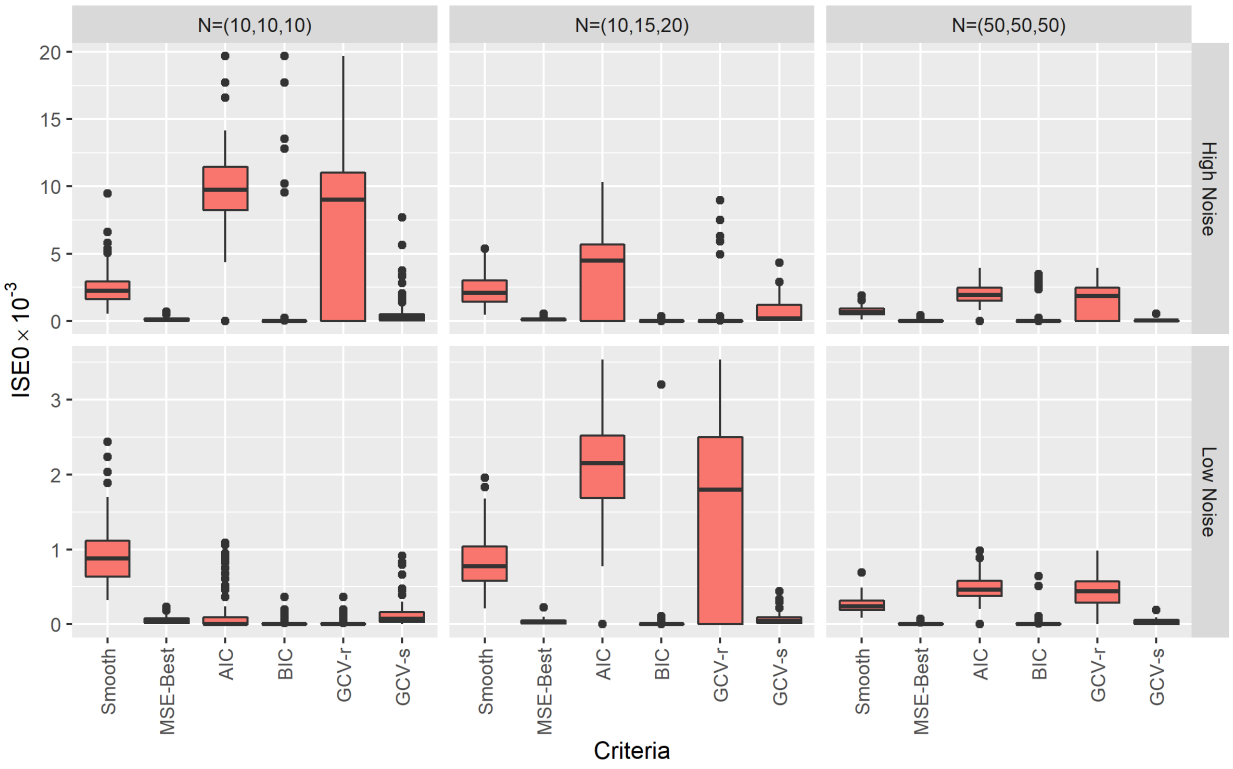


Figure 3.3: Common function case: Integrated squared error in null regions (ISE0) $\times 10^{-3}$ of the smoothed and sparsed contrast function from Gaussian noise case using 7 selection criteria

when we the fits from a random iteration in the AR(1) noise case.

Figure 3.11 looks into the fits by different criteria in an iteration of low AR(1) noise and 10 curves per group when estimated covariance matrix is used. BIC and GCV-r tend to pick high λ value which will sparse out a larger region than the true zero region. This leads to the need to compensate with low γ values for a rougher fit to match up with the data in the nonzero regions. The results are fits that are too sparse around non-zero regions but rougher in zero regions. GCV-s produces a reasonable estimates that identify relatively well sparse regions and produce nice smooth estimate over nonzero regions.

Figure 3.12 plots the fits in an interaction with the same data generation mechanism as of Figure 3.11 but the fits are done with the number of sample curves is 50 for each group. Due

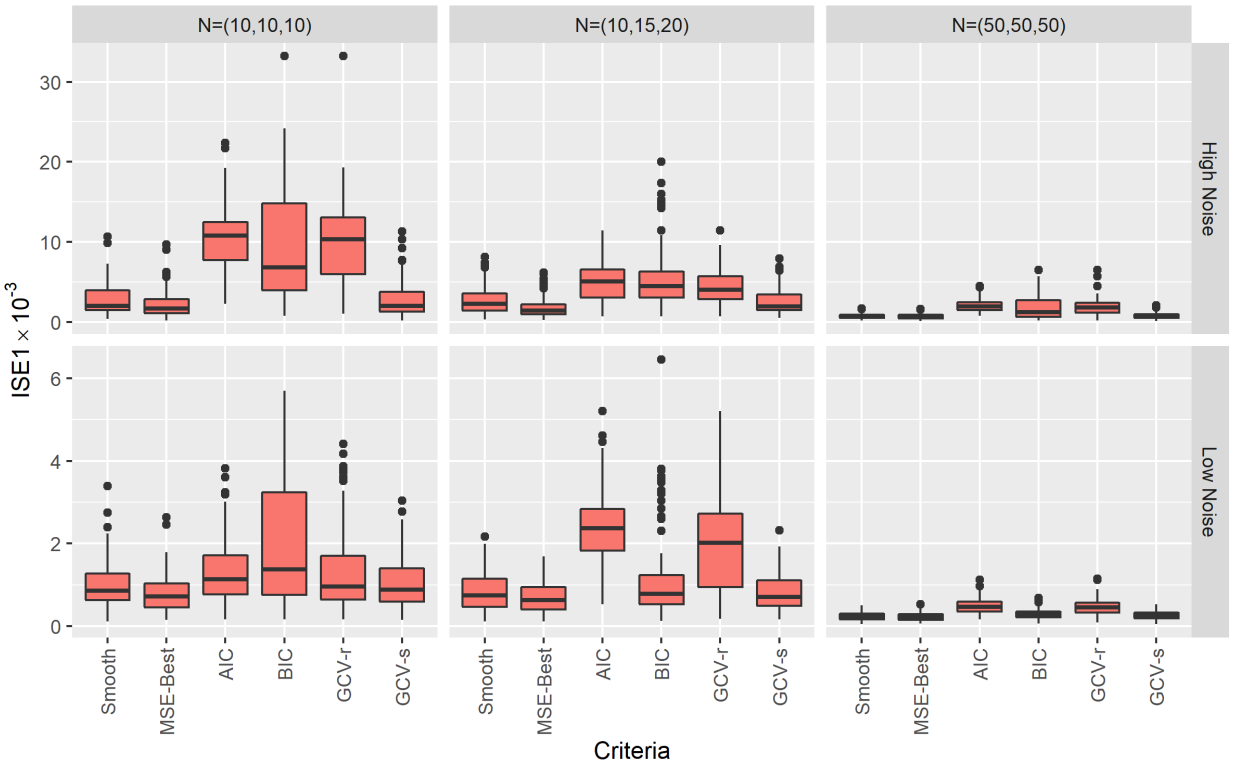


Figure 3.4: Common function case: Integrated squared error in nonnull regions (ISE1) $\times 10^{-3}$ of the smoothed and sparsed contrast function from Gaussian noise case using 7 selection criteria

to the higher sample size, all of the fits are now smoother and closer to the true contrast. However, BIC and GCV-r do not seem to produce much sparse contrast any more despite this being their strength previously. One explanation is that BIC and GCV-r values are based on the trade-off between how close the fit is the the sample contrast mean and the number of nonzero basis coefficients. In the low sample size case as in figure 3.11, BIC and GCV-r tend to pick high λ and low γ to achieve both a sparse estimate and a fit that match up very closely to the noisy sample mean. While with the high sample size case, a sparse fit may have small BIC or GCV-r values due to sparsity, however, still loses to a non-sparse fit whose RSS compared to the sample mean is much smaller due to the regression of the sample mean to a smooth and sparse true mean. On the other hand, GCV-s measures the

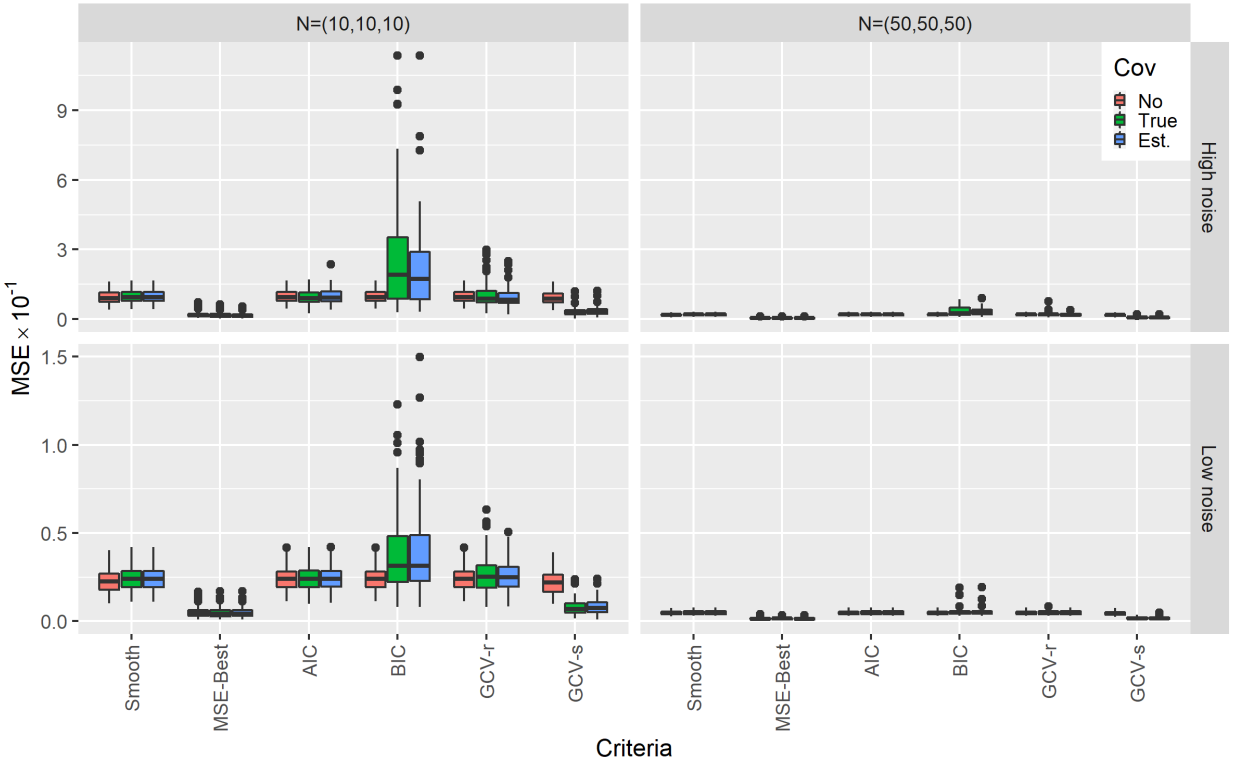


Figure 3.5: Common function case: Mean squared error (MSE) of the smoothed and sparsed contrast function from AR(1) noise case using 7 selection criteria

degree of freedom on a continuous scale, and thus, is more robust in the balance between SSE and complexity of the representation.

3.3.2 Simulation setting 2: Spectral data as observed data

We conducted a simulation study to consider the performance of our proposed method specifically for application onto contrast function from groups of spectral data. To simulate mean functions of spectral data, we take the group means from two groups of Raman spectra obtained from a dialysis session and regard them as true values from two underlying mean functions, $\mu_2(t)$ and $\mu_3(t)$. To simulate situation $\mu_1(t) = f(t) + \frac{1}{2}(\mu_2(t) + \mu_3(t))$ where $f(t)$ is the contrast function obtained from applying the contrast $(1, -\frac{1}{2}, -\frac{1}{2})$ on the three group

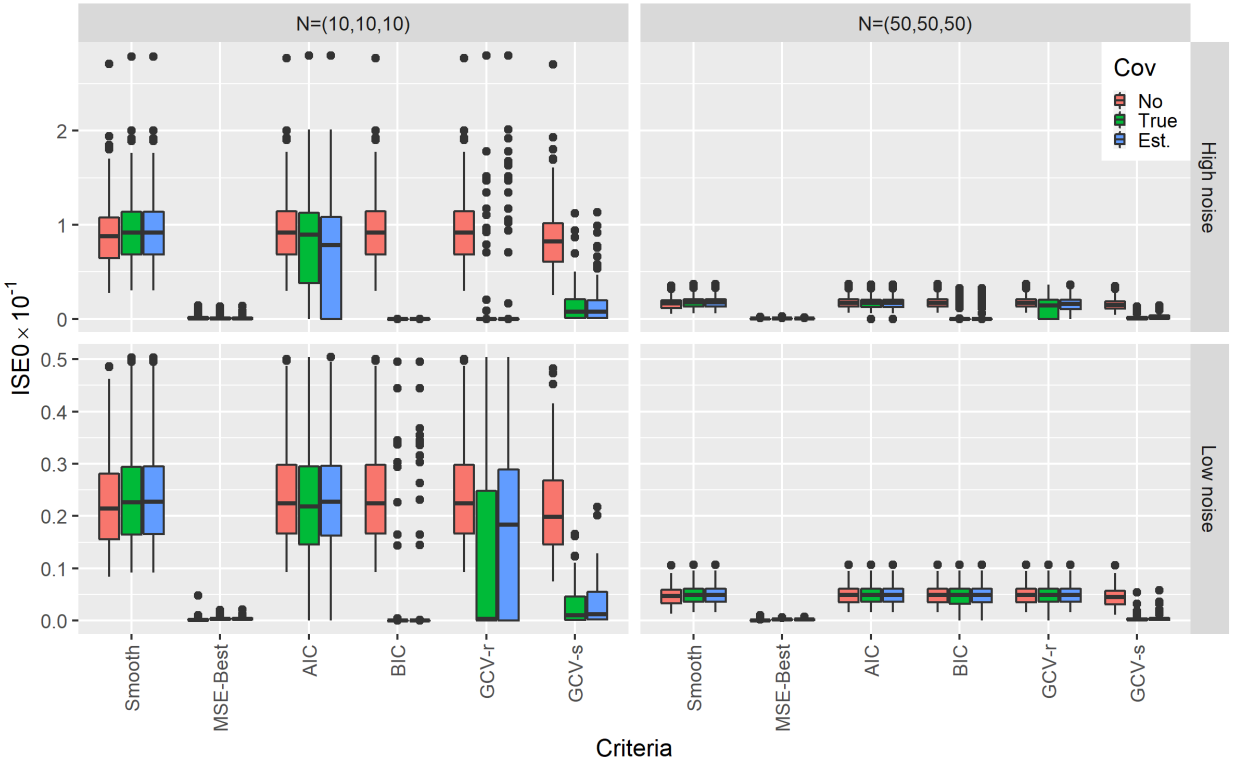


Figure 3.6: Common function case: Integrated squared error on the null regions (ISE0) of the smoothed and sparsed contrast function from AR(1) noise case using 7 selection criteria

means, we specify function $f(t)$ as follows. From the sample mean of the contrast of three groups of Raman spectra obtained from a different dialysis sample, we applied B-spline smoothing onto this function to reduce the coarseness of raw spectra data. We subset out two intervals from the output of this representation and append the remaining positions in the domain with zeroes to obtain true spectral underlying function. The resulting $f(t)$ is a sparse function that have nonzero peaks in two specified regions and is zero everywhere else on the domain.

Ultimately, this data simulation procedure produce three groups of spectral data. The spectra in each group are generated from the same group mean plus noise. The sample mean of a group, which will be labeled as group 1, is only different from the average of the sample means from the other two groups, labeled to be group 2 and group 3, at the

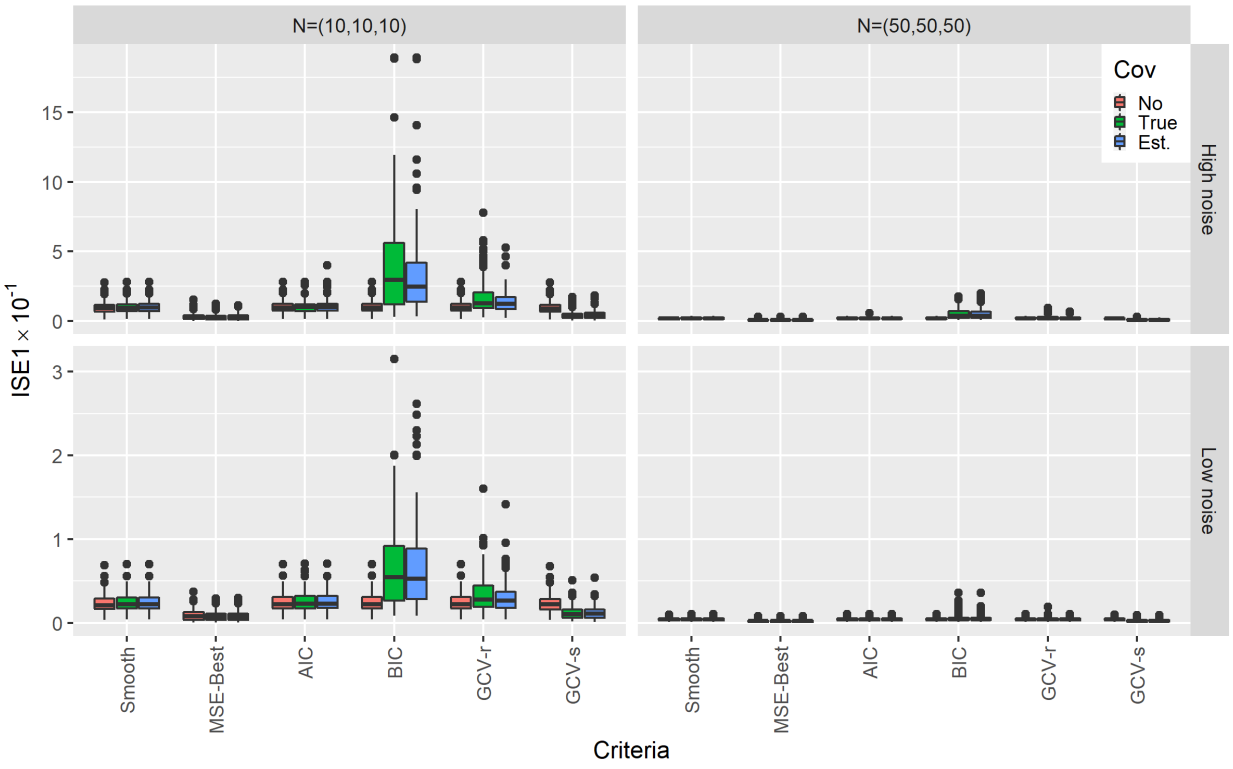


Figure 3.7: Common function case: Integrated squared error on the nonnull regions (ISE1) of the smoothed and sparsed contrast function from AR(1) noise case using 7 selection criteria

pre-determined region by an amount equal to the true function described above. The final data are discretized representation of the mean function plus noise. The curve in spectral data is of higher dimension, $n = 1024$, and domain is set to line up with the wavelength domain in Raman spectra data from hemodialysis dataset.

The three noise structures considered are Gaussian white noise, AR(1), and AR(7). Upon inspection of the dialysis data, we observed that Raman spectra possess some long-term auto-regressive dependent in the neighborhood of 7 orders. Hence, the last noise structure case is included to make the simulated data resemble the real spectral data. We consider two noise levels. Specifications for the three noise type and level are detailed in section 3.3.1.

We evaluate the method using the same performance criteria specified in section 3.3.1. In-

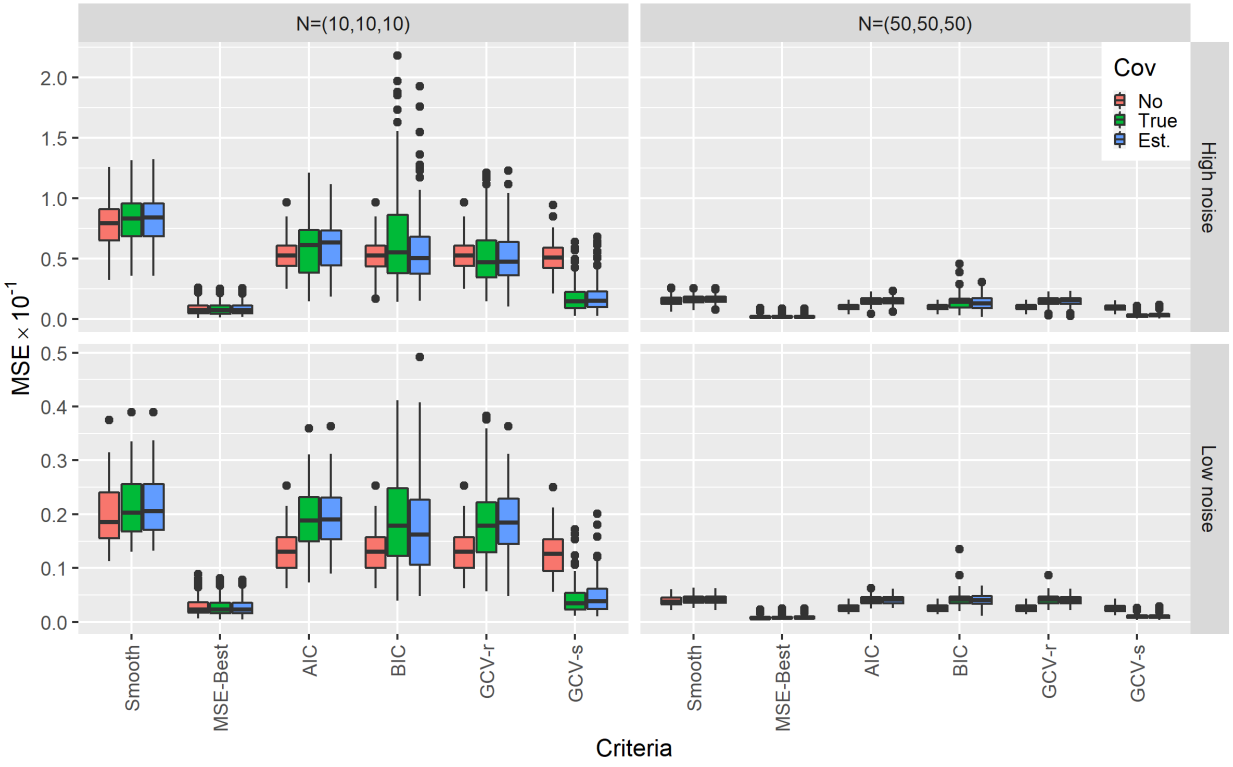


Figure 3.8: Common function case: Mean squared error (MSE) of the smoothed and sparsed contrast function from AR(7) noise case using 7 selection criteria

stead of the exact expression of the true contrast function, $f(t)$, we only have access to the exact values of this function at each wavelength on the domain. Hence, the integrals in MSE, ISE0, ISE1 criteria from 3.3.1 are replaced with a summation over the discrete values and averaged by $n = 1024$. Figures 3.13 to 3.15 display the evaluation criteria under different selection criteria. As with the case of common function, methods using sparsity regularization outperform just smoothing with roughness penalization. Tuning parameters chosen by all data-driven selection criteria lead to comparable MSE results, except for AIC with some high error replications. When the number of curves increase, errors go down for all selection methods. Compared with the common function case, we also notice some similar behaviors among the selection criteria. Specifically, BIC and GCV-r tend to go with a sparser representation than GCV-s, leading to low ISE0 and high specificity but with the

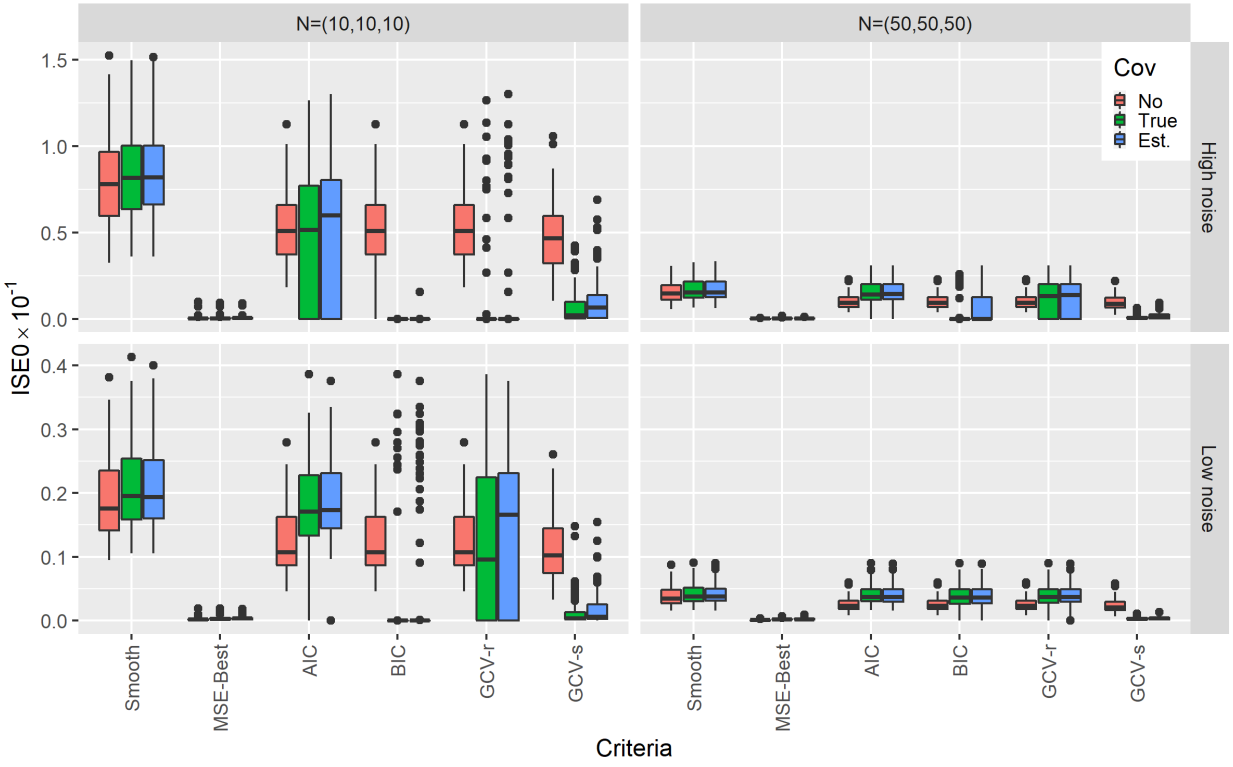


Figure 3.9: Common function case: Integrated squared error on the null regions (ISE0) of the smoothed and sparsed contrast function from AR(7) noise case using 7 selection criteria

sacrifice to ISE1 and sensitivity.

Figures 3.16 to 3.18 explore the performance of sparse contrast representation when the data is contaminated with non-Gaussian noise. The results demonstrate that using weights to account for non-Gaussian noise is an effective move. And the use of an estimated covariance matrix has comparable result to when the true covariance matrix is used.

The simulation setting for spectral data with AR(7) noise is the setting that resembles our real data. We notice that Figures 3.19 and 3.21 display similar results in terms of comparing smoothing with weight or non-weight. One noticeable result is a slight advantage in using BIC and GCV-r against GCV-s. We can attribute this to the tendency of BIC and GCV-r to favor sparsity and quickly match up to the spike signal in the nonzero region, which is

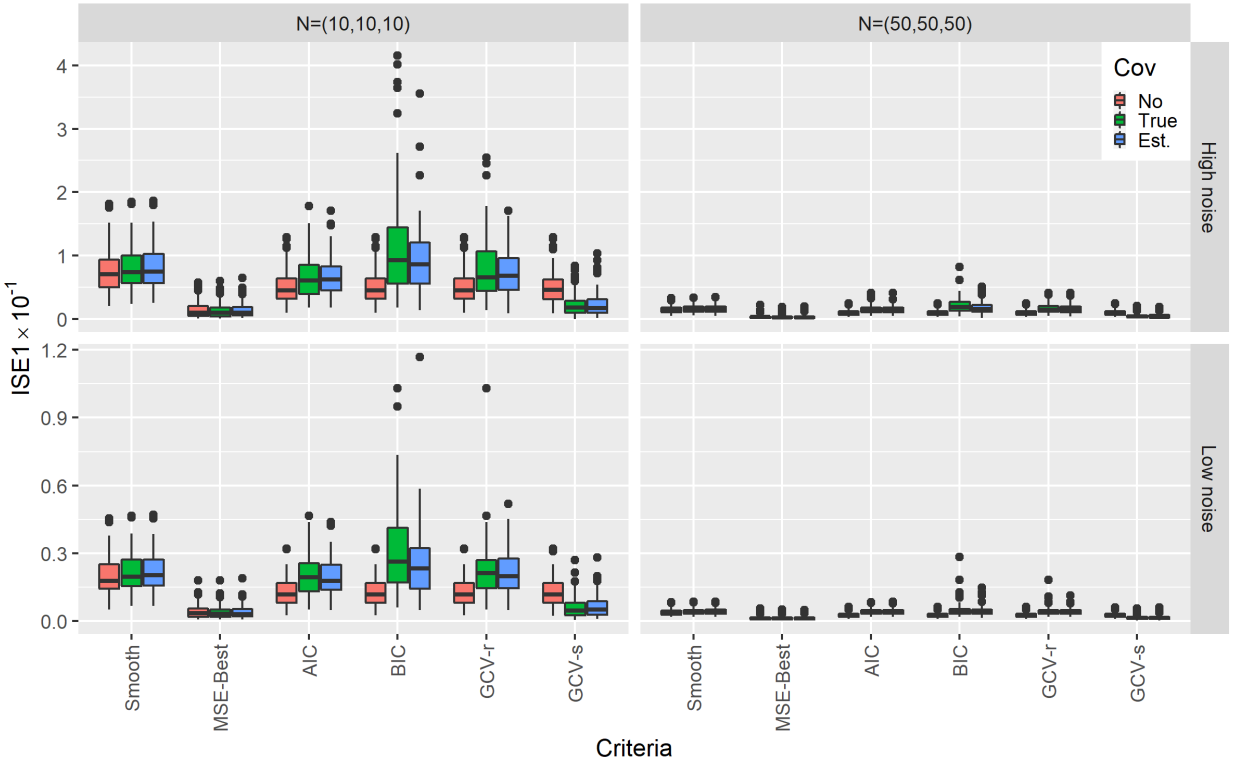


Figure 3.10: Common function case: Integrated squared error on the nonnull regions (ISE1) of the smoothed and sparsed contrast function from AR(7) noise case using 7 selection criteria

the characteristic in this spectral data.

We display the fit from one of the replications in Figure. Under spectral data, GCV-s picks up small deviations that caused by noise, while BIC and GCV-r choose setting that favor sparser result, leading to fits that recreate the true contrast well.

3.4 Application

In this section, we demonstrate the application of our proposed sparse contrast testing technique on data from real hemodialysis sessions. The study and data are conducted and collected by the lab of Dr. John Robertson and Dr. Ryan Senger from Virginia Tech. In this

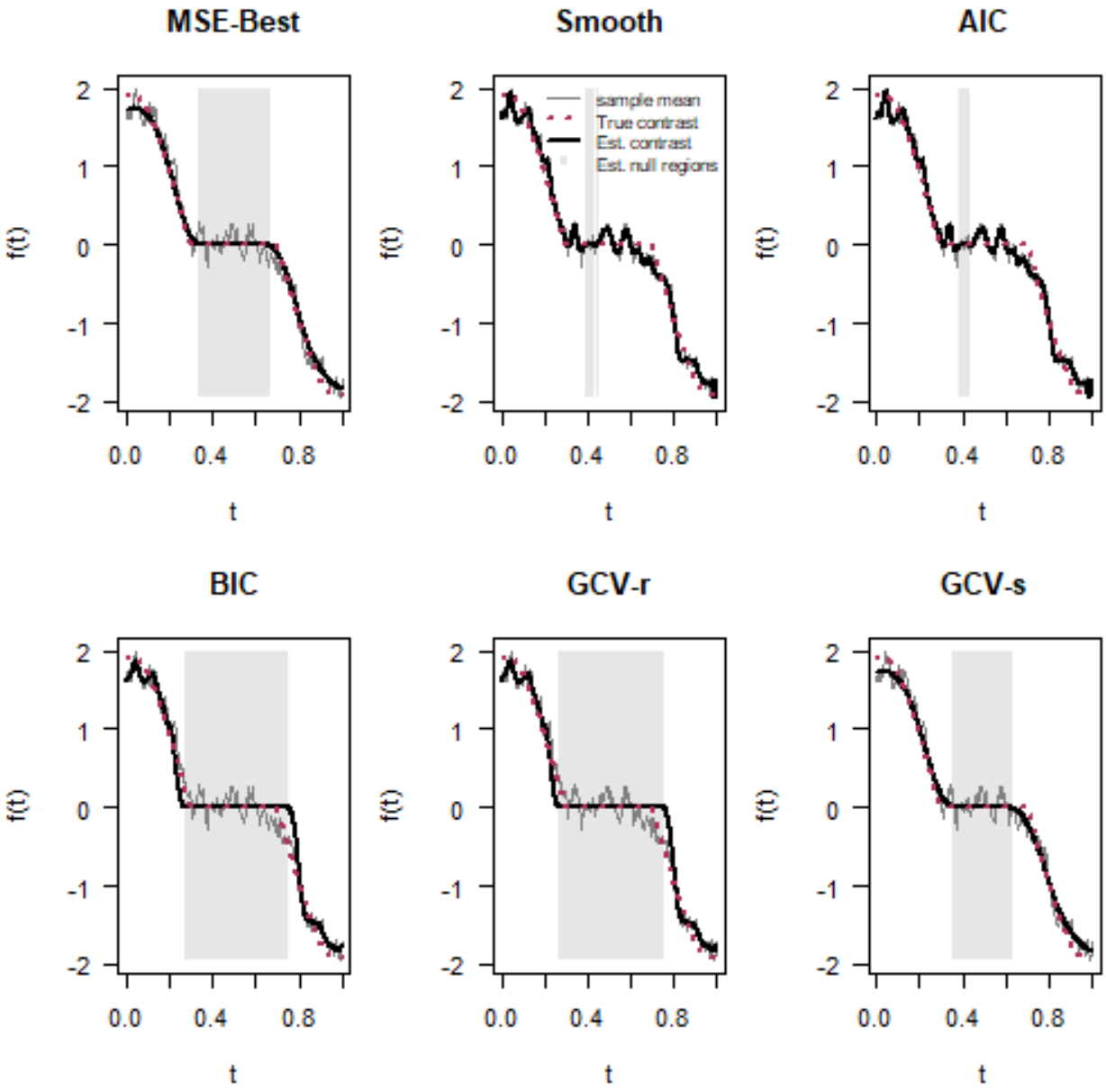


Figure 3.11: Estimates for contrast function of three groups of 10 functional data curves simulated with a common function contaminated with low AR(1) noise case

study, patients with chronic kideny diseases went through hemodialysis treatment. Each treatment session typically took 4 hours. During the session, a hemodialysis machine drew blood from patient body into a dialyzer where fresh dialysate solution cleansed metabolic waste from patient's blood. Used dialysate was pumped out and cleaned blood was returned

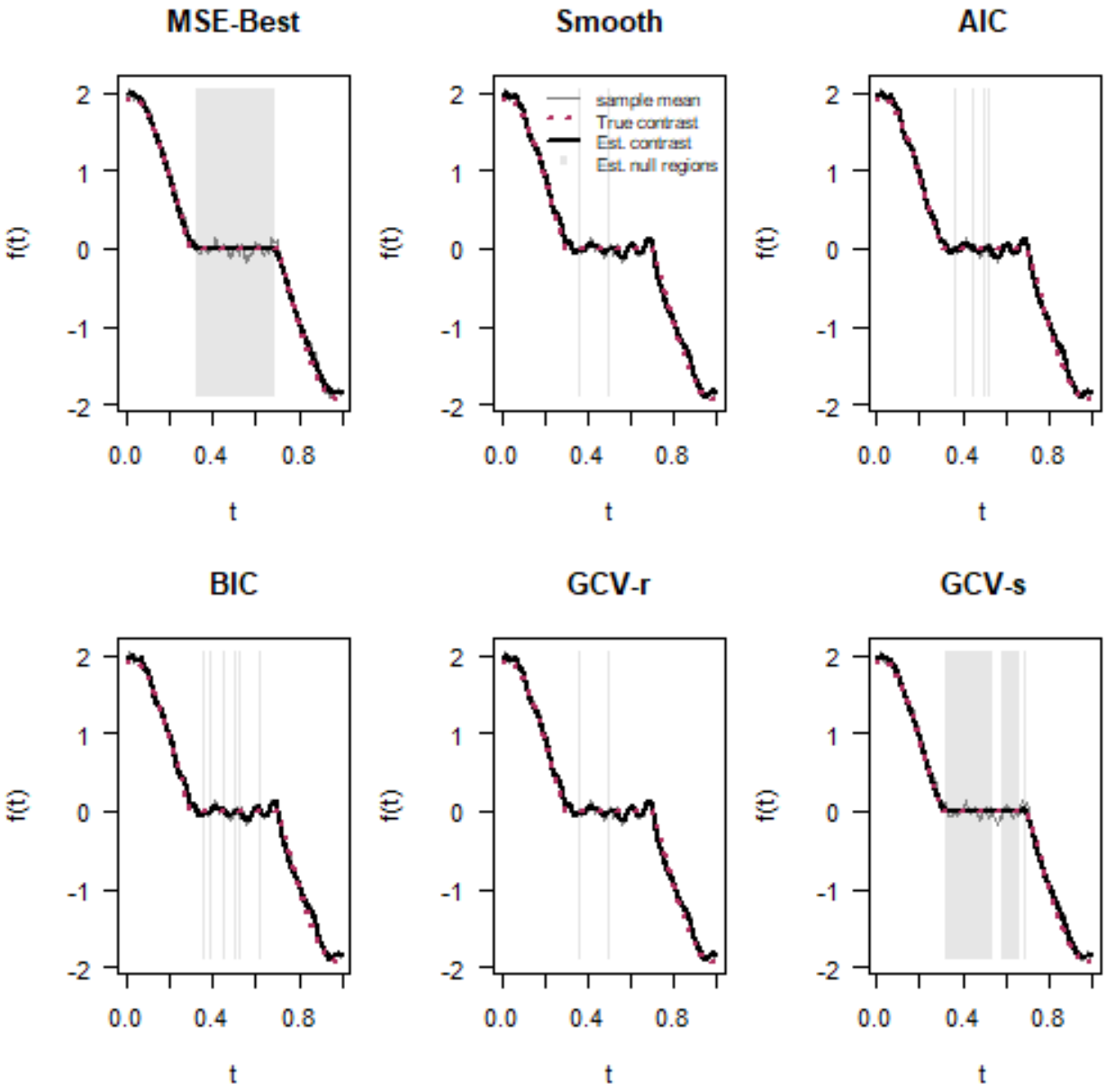


Figure 3.12: Estimates for contrast function of three groups of 50 functional data curves simulated with a common function contaminated with low AR(1) noise case

to the patient body. Normally being discarded, used dialysate solution contains chemicals that can be useful information to determine the efficacy of the treatment session in real time. Raman spectroscopy provides information on the chemical composition of the used dialysate in the form of spectra data along a dense curve. It is a mature technology and can obtain

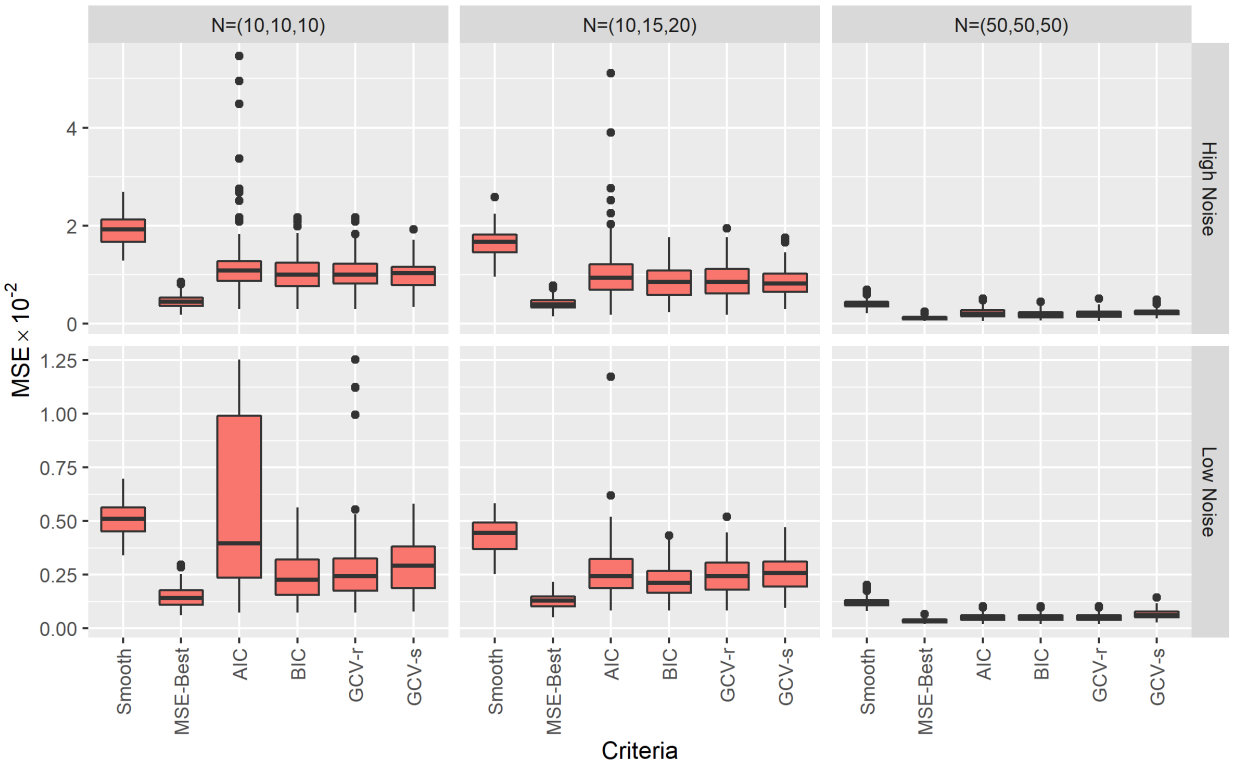


Figure 3.13: Spectral data case: Mean squared error (MSE) $\times 10^{-3}$ of the smoothed and sparsed contrast function from Gaussian noise case using 7 selection criteria

information without destructing the samples.

Our data sets are Raman spectra of used dialysate samples from real hemodialysis sessions. In each session, used dialysate samples were collected at 5 different time stamps from the start of the session: 10 minutes, 60 minutes, 120 minutes, 180 minutes, and 240 minutes. The liquid sample from each time point was divided into 10 portions. Each portion was scanned in a Raman spectroscope to produce a Raman spectrum. Therefore, a group of 10 spectra was used to represent the chemical composition in used dialysate from a time point. Prepossessing for each Raman spectrum consisted of a critical step of baseline correction to remove a baseline spectrum from the raw data. The baseline spectrum contained fluorescence or Rayleight scattering information, which are not of interest and may contaminate the interpretation of Raman spectrum if not removed. We performed this step using Iterative

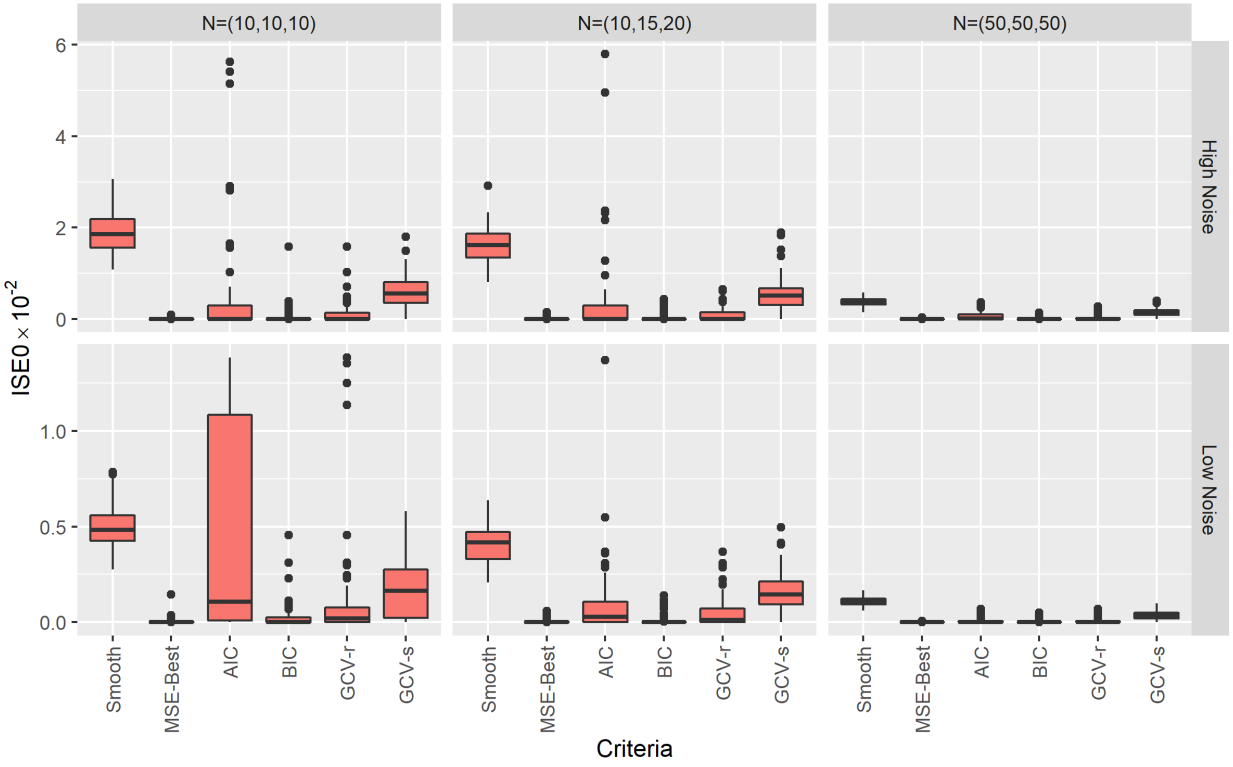


Figure 3.14: Spectral function case: Integrated squared error in null regions (ISE0) $\times 10^{-3}$ of the smoothed and sparsed contrast function from Gaussian noise case using 7 selection criteria

Smoothing-spline with Root Error Adjustment (ISREA) [71]. We applied the sparse contrast test to 3 groups of Raman spectra data for time point 1 (10 minutes time stamp), time point 4 (180 minutes), and time point 5 (240 minutes) for the contrast hypotheses

$$H_0 : \mu_1(t) = \frac{1}{2}\left(\mu_4(t) + \mu_5(t)\right) \text{ vs. } H_1 : \mu_1(t) \neq \frac{1}{2}\left(\mu_4(t) + \mu_5(t)\right) \tag{3.13}$$

This is equivalent to detect whether there exists any difference between the mean Raman spectra at the first time point and the average of the last two time points, and if so, where on the spectral domain the difference lies. The result of this step will inform the efficacy of the hemodialysis treatment session for a patient. We present the results from three different hemodialysis sessions in Figure 3.24.

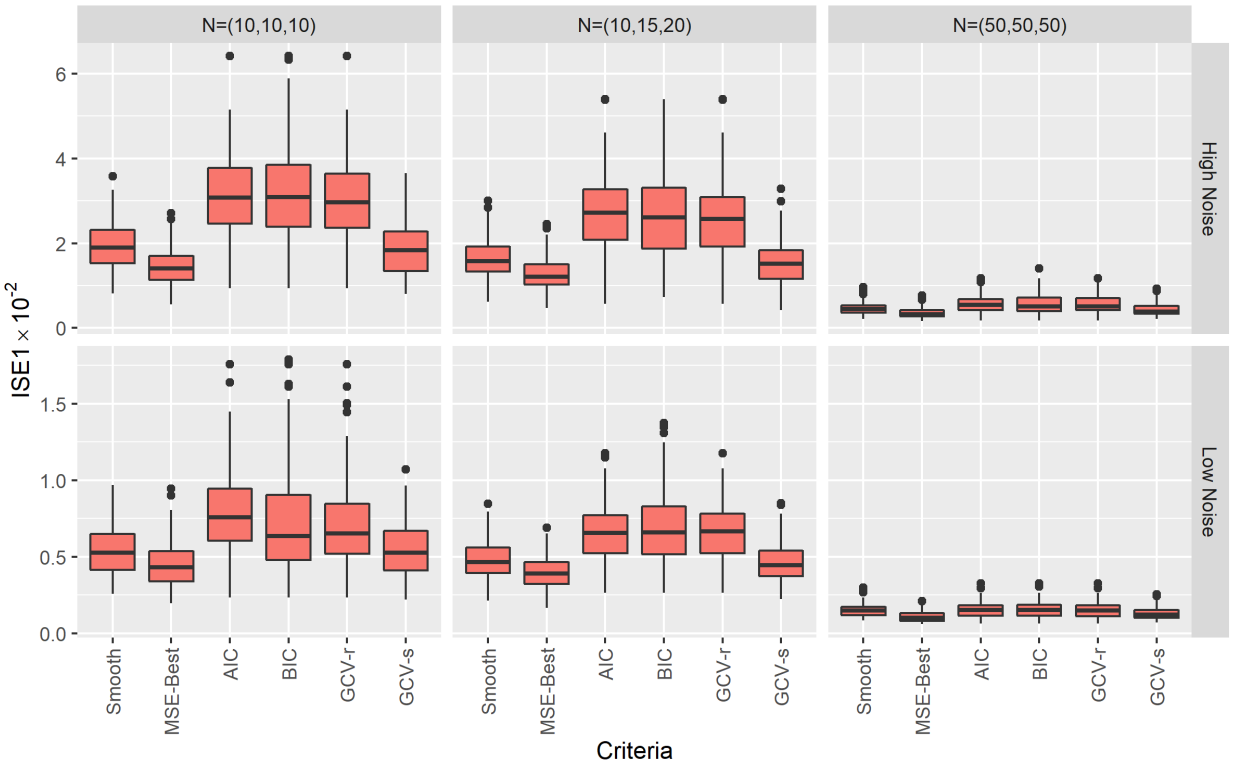


Figure 3.15: Spectral data case: Integrated squared error in nonnull regions (ISE1) $\times 10^{-3}$ of the smoothed and sparsed contrast function from Gaussian noise case using 7 selection criteria

We first apply the contrast test on the data for each patient for the hypothesis (3.13). The test results across two test statistics proposed in (2.12) are consistent with each other for all three patients. The hypothesis is rejected for patient 1 ($p_1 = 1e-04, p_2 = 0$), and for patient 2 ($p_1 = 0.0085, p_2 = 0.00043$). The test did not reject the hypothesis for patient 3 ($p_1 = 0.62, p_2 = 0.60$). Using our sparse interval detection technique, a sparse estimate is obtained for each patient. We use the BIC to select the tuning parameters since the criterion performs pretty well in the spectral data simulation. The sparse representations of the contrast function detect null regions and give estimation for region of difference. Based on this estimate, we can identify chemicals with peaks falling within nonzero regions whose intensities reduced within the used dialysate from time point 1 to the later two time points.

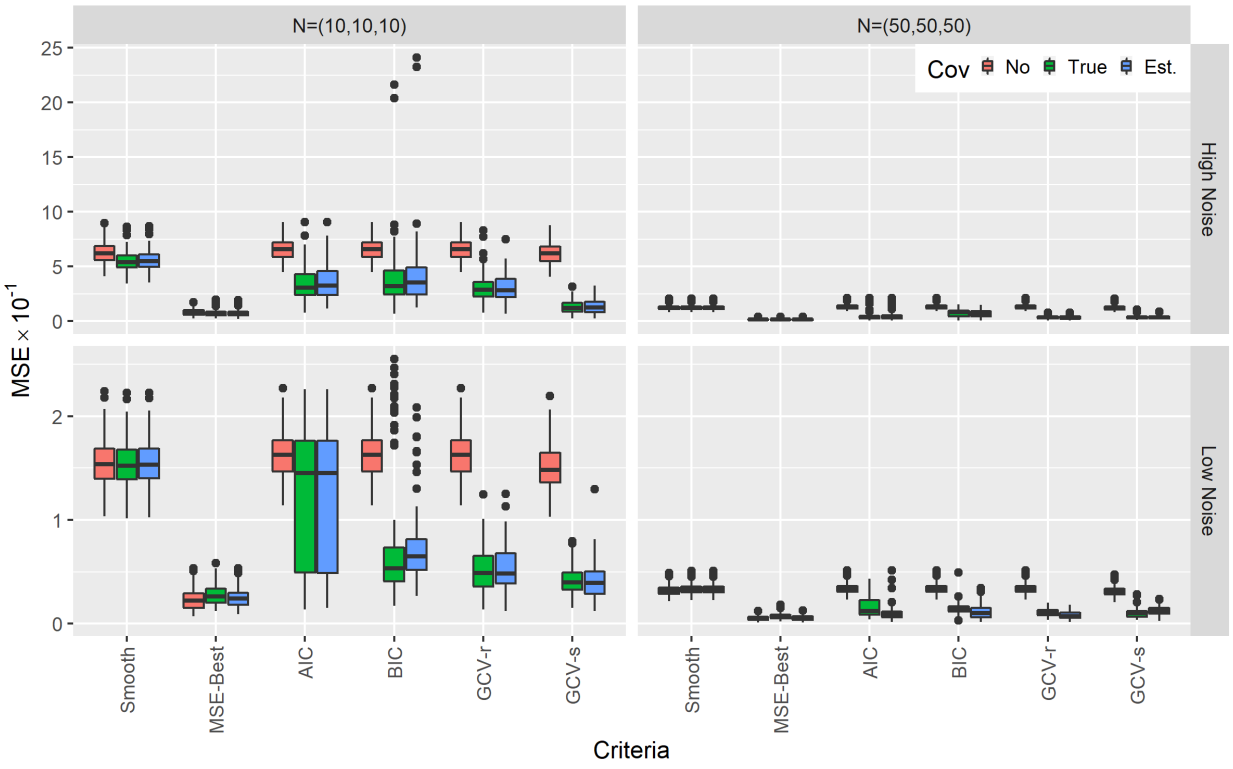


Figure 3.16: Spectral data case: Mean squared error (MSE) of the smoothed and sparsed contrast function from AR(1) noise case using 7 selection criteria

3.5 Conclusion

In this work, we propose a method to obtain a locally sparse representation of the contrast function under a functional ANOVA model. When the functional ANOVA test and followed-up contrast test is rejected, our method provides a local sparse representation, which could help practitioners pinpoint the subregion(s) of the domain where the difference of functional data groups exists. We expanded the standard roughness penalty smoothing by adding an L_1 penalty term to the objective function to obtain a sparse representation. The estimation is through a B-spline expansion and the final presentation using B-spline basis is a sparse function with smooth estimates over nonzero regions, enhancing the interpretation ability of the contrast test. We also examine a variety of tuning parameter criteria.

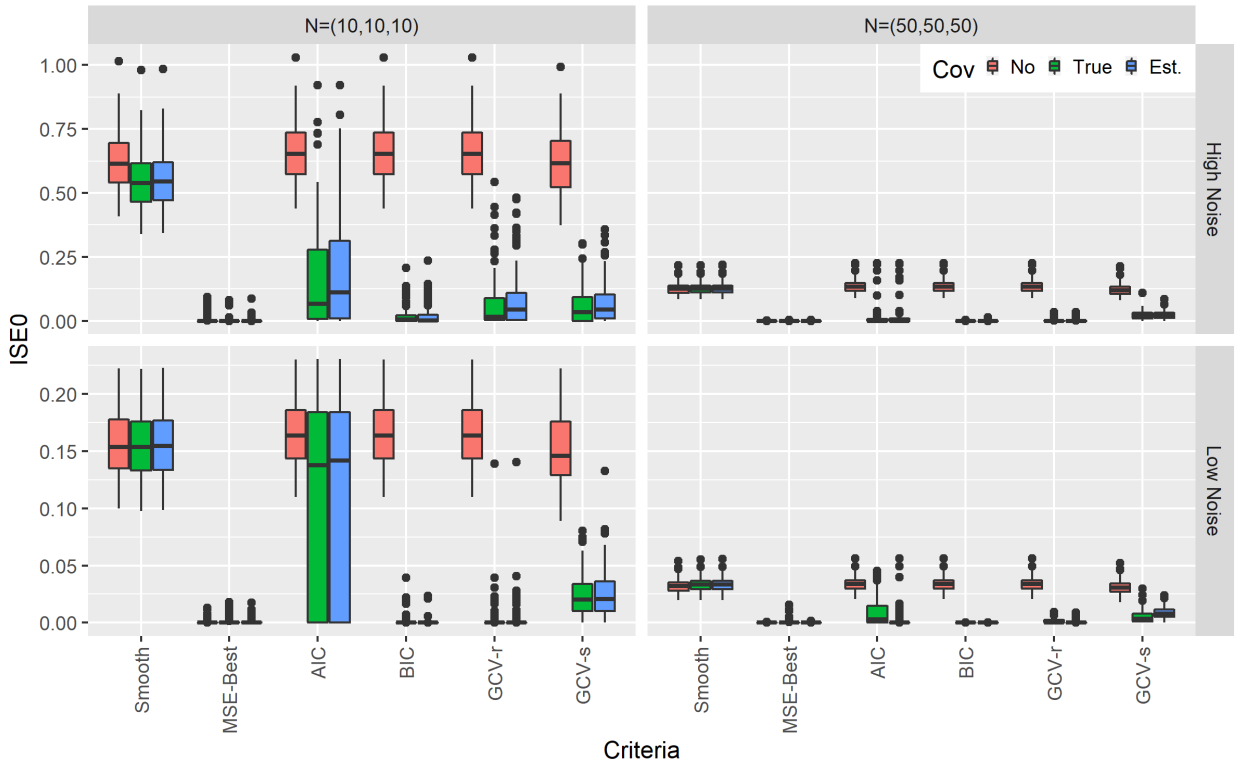


Figure 3.17: Spectral data case: Integrated squared error on the null regions (ISE0) of the smoothed and sparsed contrast function from AR(1) noise case using 7 selection criteria

Our extensive simulations demonstrate the benefit of the sparse representation of a contrast function. With the additional functional SCAD penalty, the sparse representation outperforms the standard smooth only representation. GCV with a continuous degree of freedom and some correction is showed to effectively tune smoothing and sparsity parameters in the case when the underlying function is smooth and sparse. When the underlying function is rougher and spectrum-like, the BIC and GCV using a discrete degree of freedoms stand out to be the best tuning criteria. We applied the procedure to the monitoring of hemodialysis adequacy via the comparisons of Raman spectra generated from dialysate samples collected at different times points of the treatment session. The nonzero regions identified in the study provide useful information on critical chemicals that play a key role in hemodialysis monitoring.

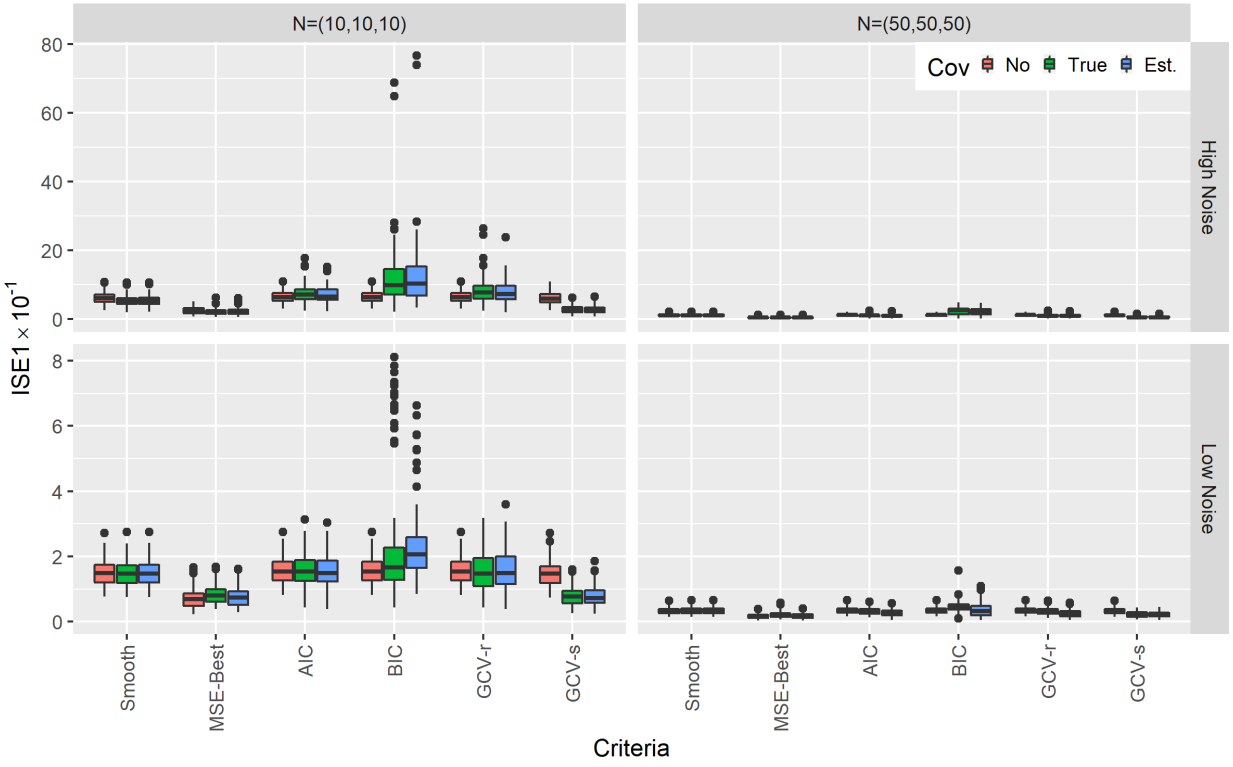


Figure 3.18: Spectral data case: Integrated squared error on the nonnull regions (ISE1) of the smoothed and sparsed contrast function from AR(1) noise case using 7 selection criteria

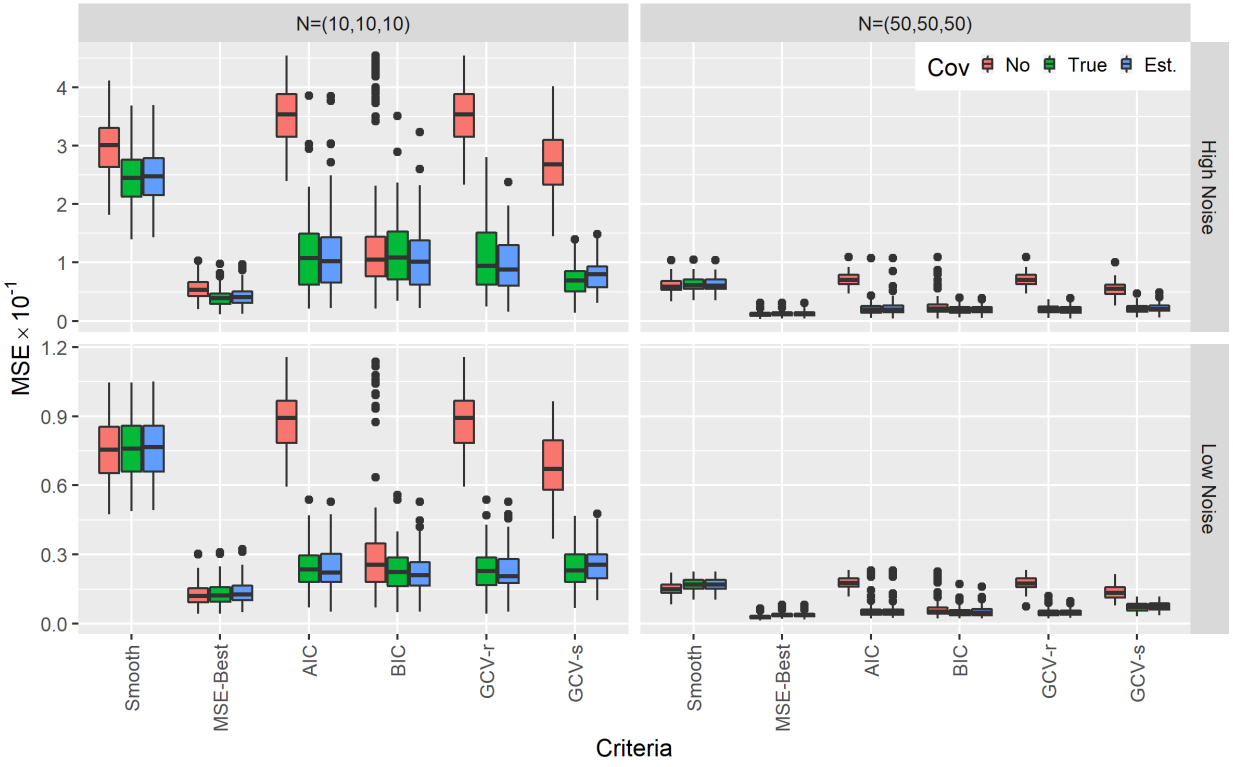


Figure 3.19: Spectral data case: Mean squared error (MSE) of the smoothed and sparsed contrast function from AR(7) noise case using 7 selection criteria

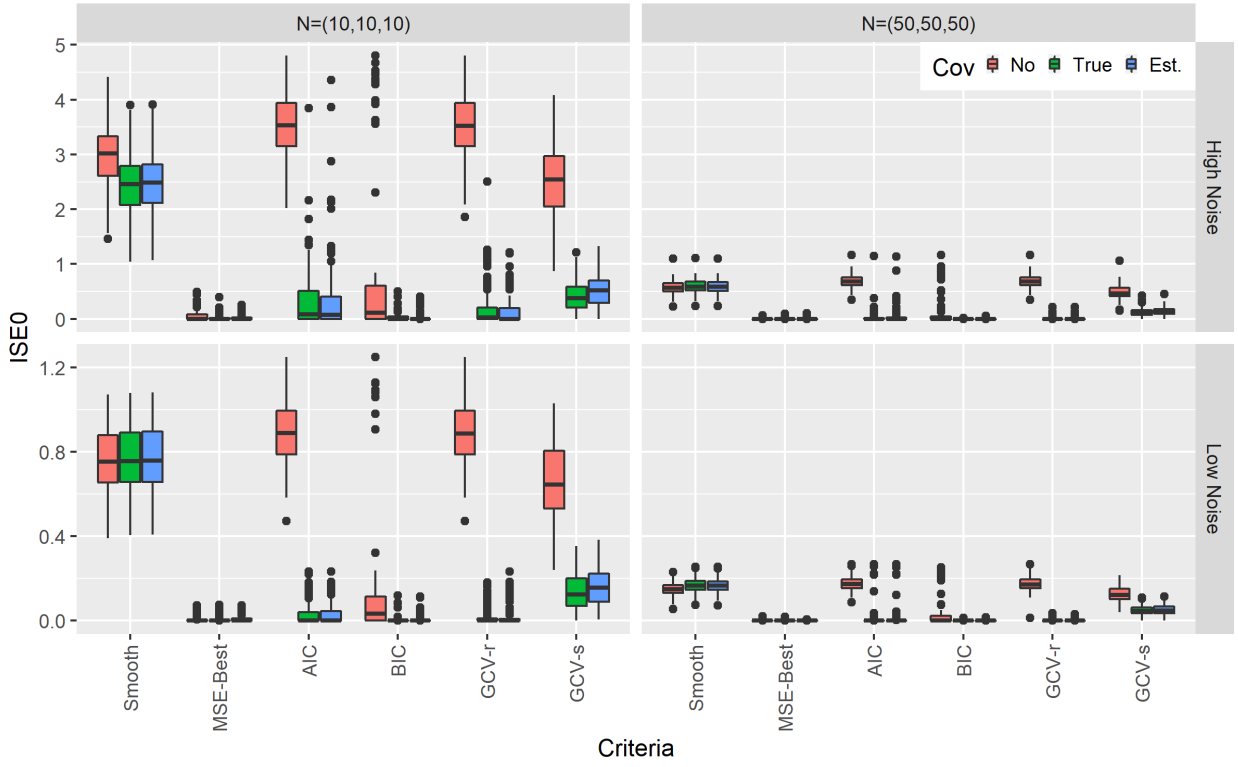


Figure 3.20: Spectral data case: Integrated squared error on the null regions (ISE0) of the smoothed and sparsed contrast function from AR(7) noise case using 7 selection criteria

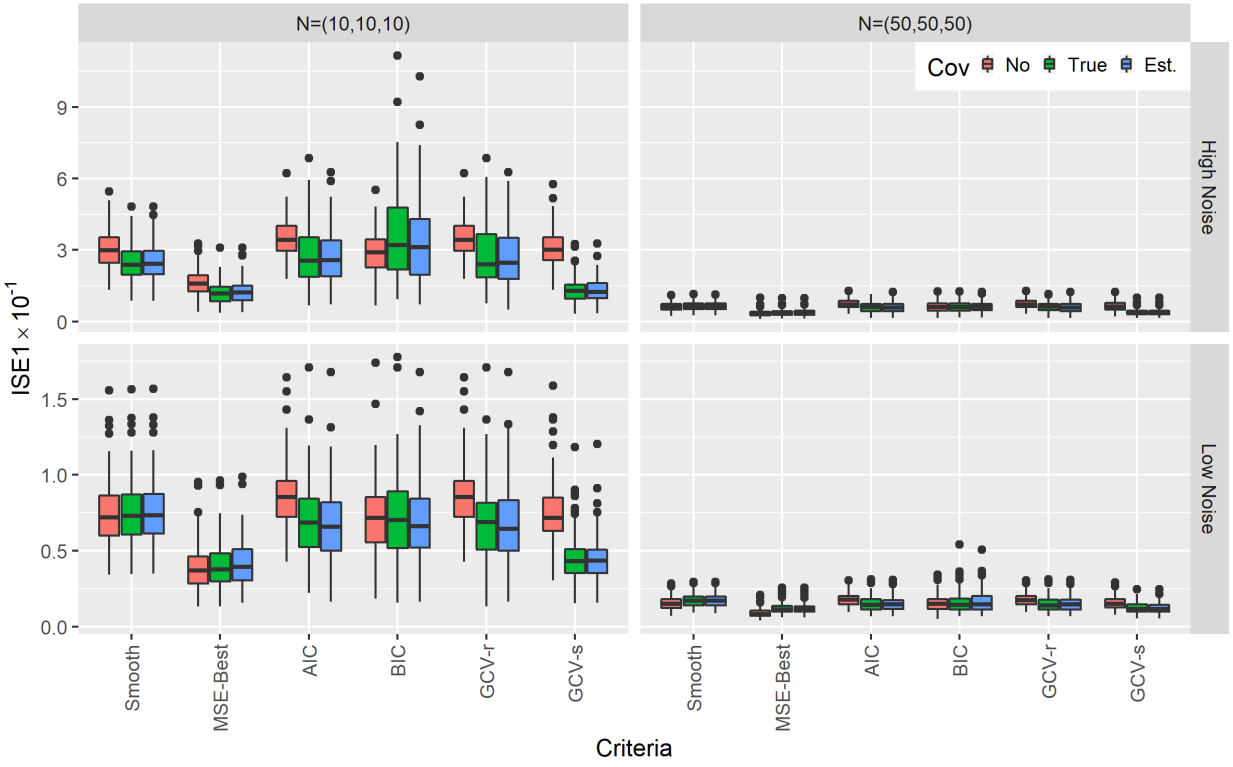


Figure 3.21: Spectral data case: Integrated squared error on the nonnull regions (ISE1) of the smoothed and sparsed contrast function from AR(7) noise case using 7 selection criteria

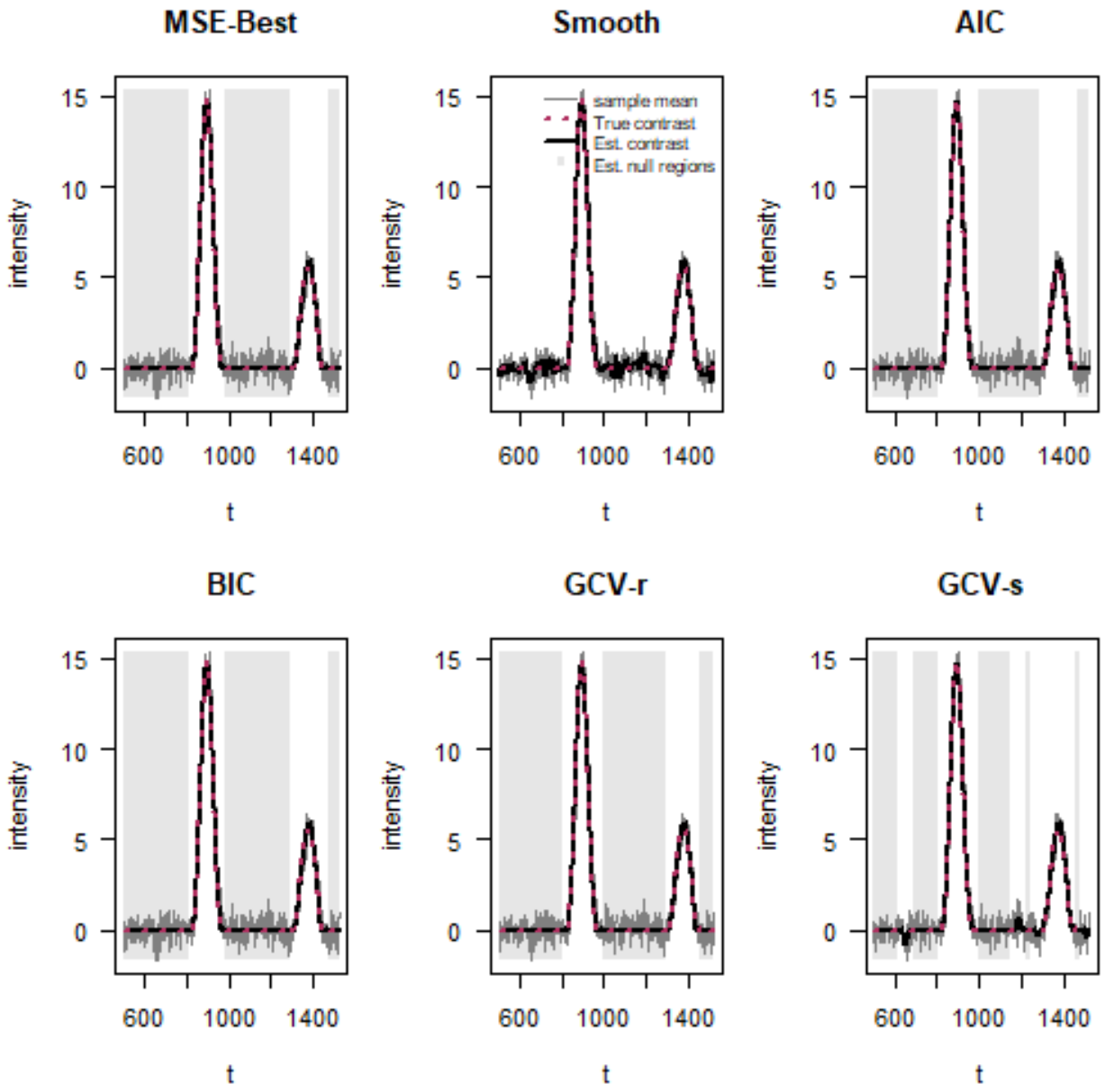


Figure 3.22: Estimates for contrast function of three groups of 10 functional data curves simulated with spectral data contaminated with low AR(7) noise case

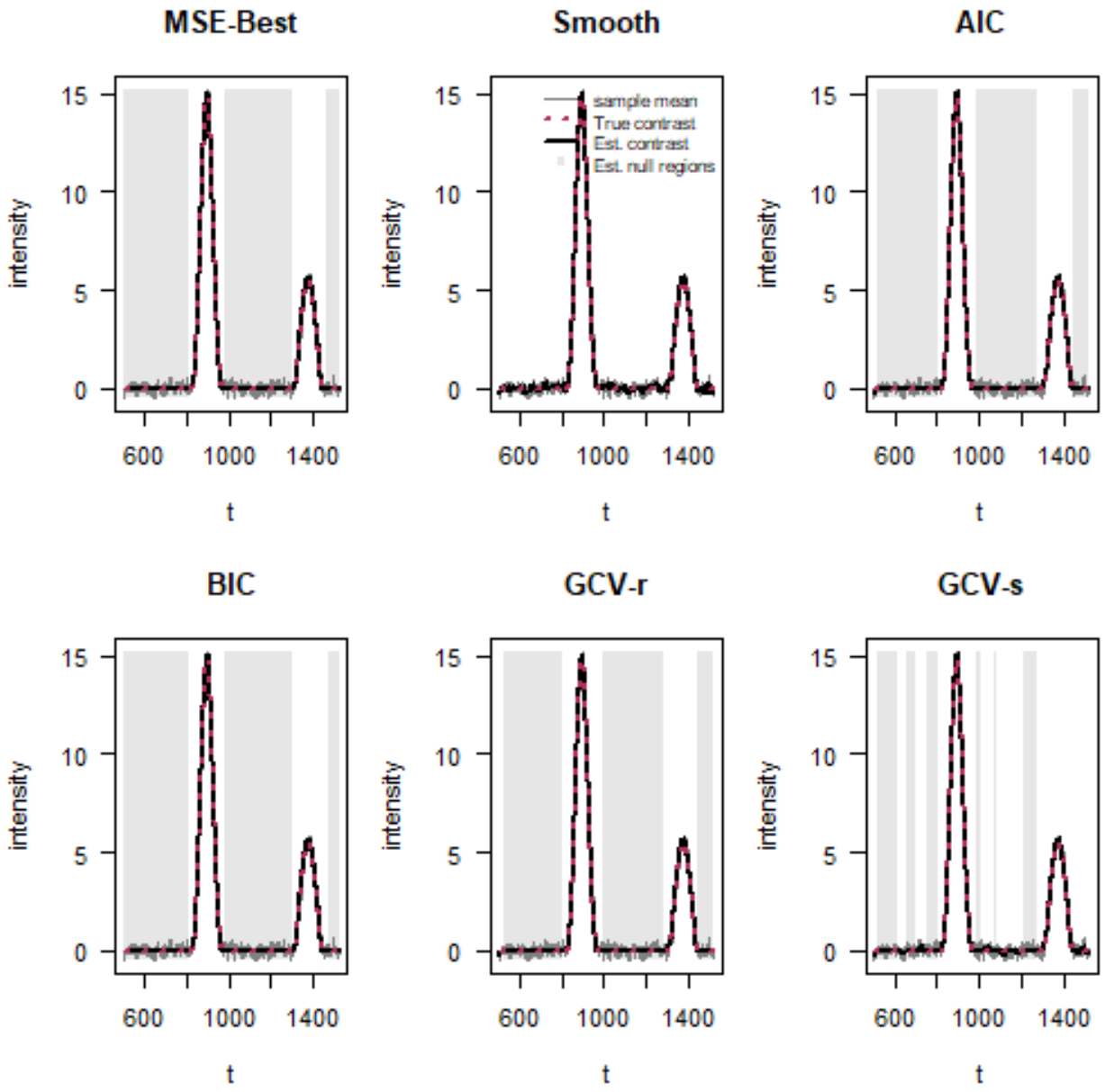


Figure 3.23: Estimates for contrast function of three groups of 50 functional data curves simulated with spectral data contaminated with low AR(7) noise case

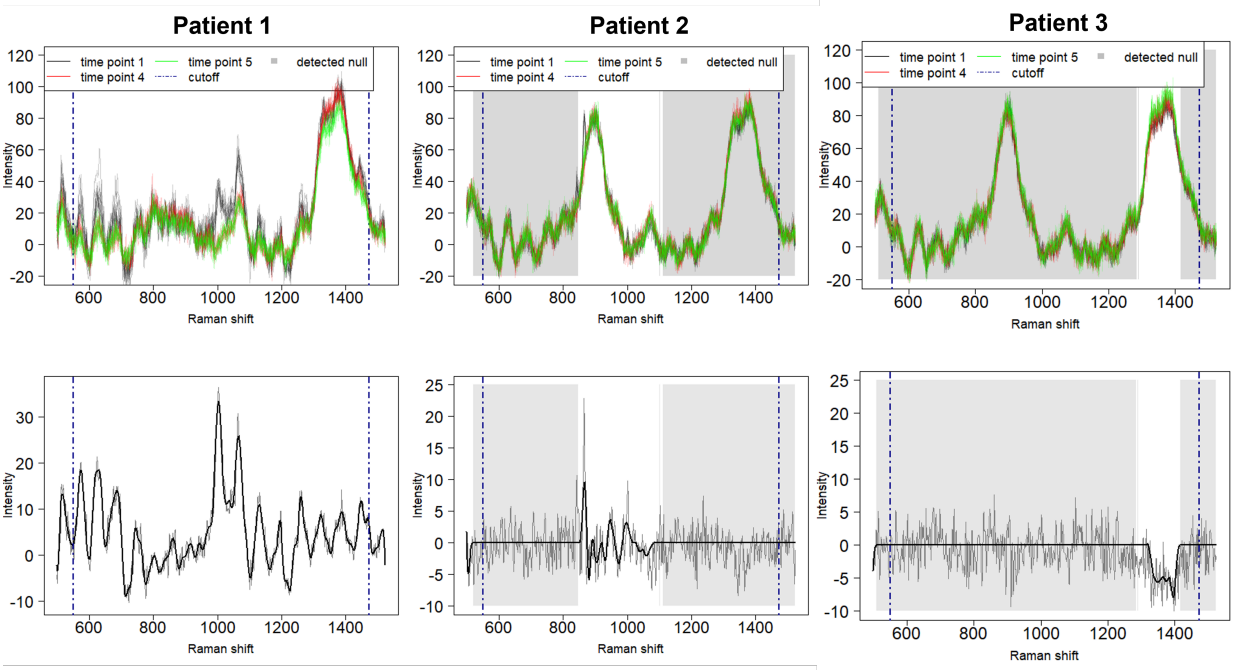


Figure 3.24: Sparse contrast testing result of the contrast $\mu_1(t) - \frac{1}{2}(\mu_4(t) + \mu_5(t))$ of Raman spectra data for three different patients (column). Top row plots the data by time points. Bottom row plots the estimated sparse contrast function. Gray shade mark regions where contrast function is estimated to be zero.

Chapter 4

Reliability Study of Battery Lives: A Functional Degradation Analysis Approach

4.1 Introduction

Renewable energy plays an important role in global energy security and in the controlling of global warming [40]. Commonly used renewable energy technologies include solar power technology, wind turbines, biomass technology for power generation and tidal energy technologies (Bull [4] and Panwar et al. [41]). Once electricity is generated from those renewable energy technologies, energy storage batteries are needed for energy to be stored when available and released to when needed. Li-ion batteries are widely used for energy storage and one significant application is on electric vehicles (Diouf and Pode [10] and Castelvecchi et al. [7]).

An important question in these applications is the assessment of battery life under a certain use condition. An accurate assessment of battery lives is critical for the large-scale deployment of renewable technology. This motivates battery life studies like the one reported in Saha and Goebel [54], where the question of interest is how many charge-discharge cycles a

battery can have before it reaches a failure threshold in terms of capacity loss.

Even though there is vast literature on the modeling and prediction of Li-ion battery lives, existing analyses are mostly straightforward applications of traditional statistical methods from a simplistic view of the data. For example, existing studies do not take full advantage of the whole discharge curves, whereas each curve gives much more details, than a simple summary number such as the area under the curve, on how the battery performed in a cycle. Figure 4.1 illustrates the available voltage discharge curves (VDCs) of two batteries in the data set of interest. An aggregation summary such as area under the curve does represent a type of degradation measure, but there is clearly the risk of information loss since the whole curve is not taken into account. This motivates us to adopt the functional data analysis (FDA) approach to model and predict battery lives. Under the FDA framework, each curve is viewed as a functional observation and we shall model the VDCs directly through mixed effects models. Ever since its introduction by the seminal monograph Ramsay and Silverman [49], FDA has found many extensions to various areas in statistics; see, e.g., the recent reviews in Wang et al. [69], Aneiros et al. [1], and Li et al. [23]. To the best of our knowledge, our work is the first FDA approach in the area of battery life modeling, and more generally, in the area of degradation modeling.

Modeling functional curves with covariate information falls under the umbrella of functional response models. Existing functional response models require a common domain for the functional responses. However, in our battery data, each VDC has its distinctive domain, ranging from the start of the discharge cycle to the so-called *end of discharge (EOD)* time [55]. Therefore, we first introduce a scaled version of VDCs where each VDC is scaled by its EOD to lie on the common domain of $[0, 1]$. Then we consider a model consisting of two parts. One part models the EODs and the other models the scaled VDCs. For the predictors in the models, experimental conditions are the natural covariates to include. The

cycle number is also necessary for the modeling of the degradation patterns. Some other covariates in consideration include the resting time between cycles, the EOD of the previous cycle, and the scaled VDC itself. To model the scaled VDCs, we first conduct a functional principal component analysis to represent each curve by its projected scores into the leading functional principal components. Then a multivariate mixed effects model is fitted with the random effects introduced to incorporate the within-battery correlations between VDCs. For the EODs, we consider two version of mixed effects models, with or without the inclusion of the functional covariate of scaled VDC. Our degradation analyses, with three versions, are all based on these functional models of the scaled curves and EODs. The first version is the standard general path model where degradation amounts computed from fitted VDCs at the observed cycles (or training cycles) are the only inputs to fit the prediction model for degradation amounts. The other two versions, corresponding to the two versions of EOD models, would use the functional models to predict scaled VDCs and EODs first and then use them to calculate the predicted degradation amounts.

Through extensive simulations and comprehensive analysis of the battery data, we show the great potential of a fully FDA approach to degradation analysis of batteries. In particular, it has the following advantages against the standard or existing approaches: (i) the framework allows an easy incorporation of covariates such as experimental conditions, which are critical factors for battery degradation but have been largely ignored by the existing literature; (ii) it makes full use of the whole VDCs, ensuring a minimal loss of information from the original data; (iii) it is demonstrated in our numerical studies to have much better prediction performance than the traditional aggregration methods, which is essential in degradation analysis; and (iv) its predictions include the predictions for the EODs and scaled VDCs, which can be assembled to obtain a prediction for the original VDC at a future cycle.

The rest of the paper is as follows. In Section 4.2, we introduce the NASA battery dataset

for the analysis. We give an overview of the existing liturature studying this dataset under battery degradation and health monitoring framework. We discuss insights we gain from these work as well as their gaps in modeling VDCs that inspire this work. In Section 4.3, we present the methodology in details including notations, modeling, predicting and degradation analysis. In Section 4.4, we conduct a simulation study to test the performance of our proposed method against a standard method when we hold the truth about the data generation mechanism. In Section 4.5, we present the real data analysis using the developed framework. Finally, we conclude with some summaries and areas for future research in Section 4.6.

4.2 Battery Life Study from NASA Ames Prognostics Data Repository

The study we consider here is from the NASA Prognostics Data Repository [54]. In the study, Li-ion batteries were tested through cycles of charge and discharge at different room temperature and conditions. Experimenters performed the same charging procedure for all batteries after each discharge cycle. The lifetime measure for batteries was the number of discharge cycles. The experimental conditions for discharge cycles varied among batteries. We focus on the following conditions.

- Testing temperature (*temp*): room temperature (24°C), elevated temperature (43°C), or low temperature (4°C).
- Discharge current (*dc*): 1 Amp, 2 Amps, or 4 Amps.
- Level of voltage where discharge ends or stopping voltage (*sv*): 2 Volts, 2.2 Volts, 2.5 Volts, or 2.7 Volts.

The data were collected from 20 batteries with constant experimental conditions across discharge cycles in the experiment. Throughout the discharge cycles of each battery, measurements of discharge voltage levels were recorded. As shown in Figure 4.1 for two batteries, the VDCs for each battery clearly show a degrading pattern. In early discharge cycles, the battery voltage stays high for a longer time until a sharp drop at a later time in the cycle. As the cycle number increases, the battery voltage sustains for a shorter time and drops sharply at an earlier time in the cycle. This pattern is revealed by the motion of the discharge curves moving inward in Figure 4.1.

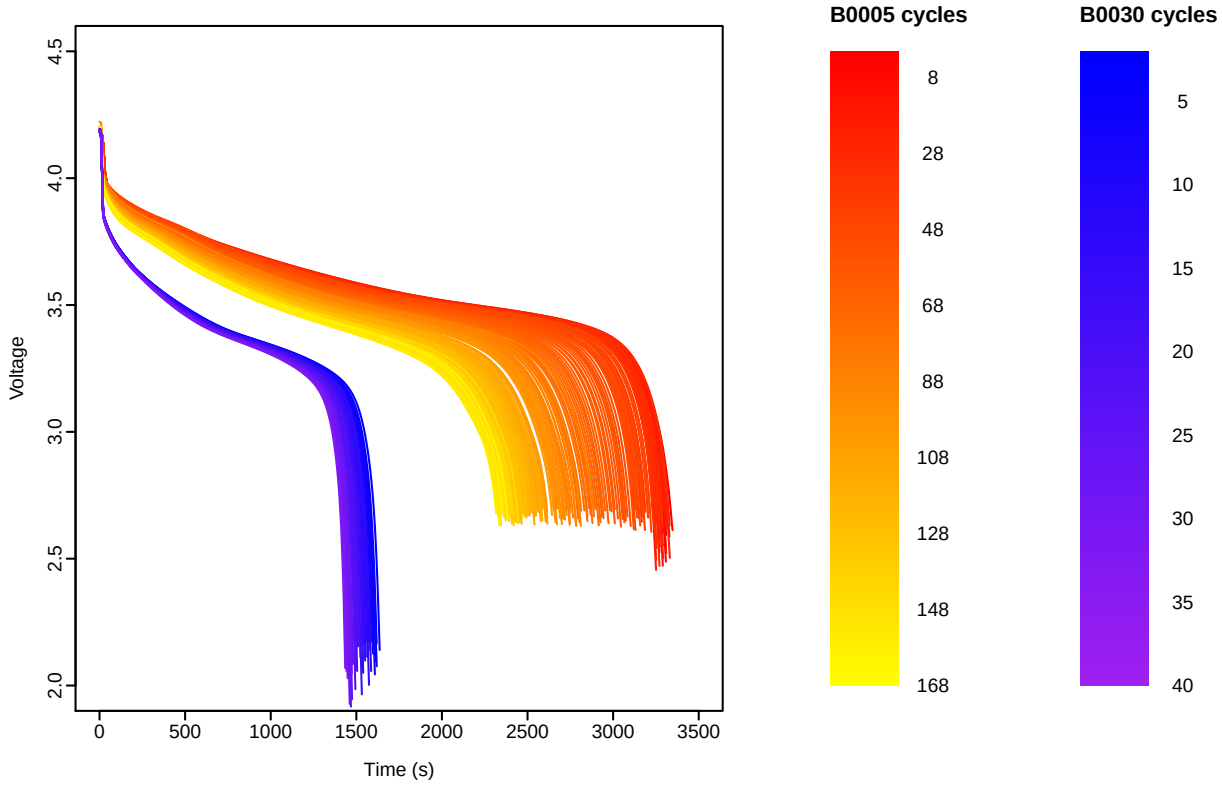


Figure 4.1: Voltage discharge curves by cycles number for battery B0005 (discharge current: 2 Amps, Stopping voltage: 2.7 Volts, temperature: 24°C) and battery B0030 (discharge current: 4 Amps, Stopping voltage: 2.2 Volts, temperature: 43°C).

This is a well-known dataset in the literature of battery health prognostics. The main focus of health prognostics is on studying the state of health (SOH) of batteries with the goal of

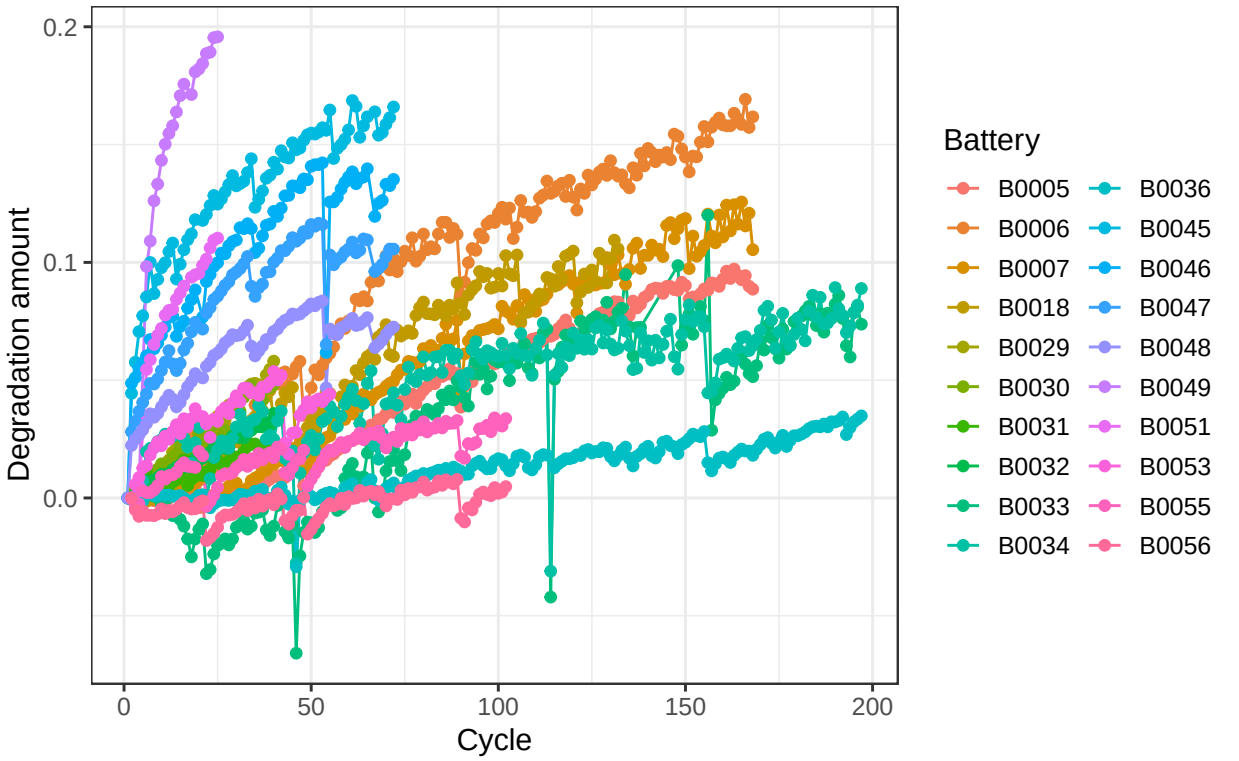


Figure 4.2: Degradation amount by cycle number and battery determined by L^1-norm of batteries' voltage discharge curves.

predicting the remaining useful life (RUL). The RUL is determined by how close a battery's capacity is to a preset threshold. As a standard measurement, the battery capacity at a cycle is closely related to the area under the corresponding VDC. Therefore, most existing methods have focused on direct modeling and predicting battery capacity along discharge cycles on the degradation path.

Among the earliest work, Saha and Goebel [55] proposed to predict battery capacity under the particle filtering (PF) framework. They also developed a Rao-Blackwellized PF approach in Saha et al. [56] to reduce uncertainty on capacity prediction. A comparison of their methods is available in Saha et al. [57]. Since then various statistical and machine learning methods have been proposed for the battery capacity prediction problem. Examples of statistical models include stochastic models through the use of autoregressive processes [27]

or Wiener processes (Tang et al. [65], Shen et al. [63]), and Gaussian process regression (Liu et al. [28], Liu et al. [29], He et al. [17], Richardson et al. [53], Tagade et al. [64]). Examples of machine learning techniques include naïve Bayes [37], support vector machine [43], relevance vector machine [31], genetic algorithms [73], and neural networks (Sbarufatti et al. [59], Nascimento et al. [36]). Liu et al. [30] provided a fusion method integrating regularized particle filtering and autoregressive models. A comprehensive review of these methods can be found in Martin et al. [33].

However, none of the aforementioned work included all the 20 batteries in their analyses. Most work focuses only on the first four batteries (B0005, B0006, B0007, and B0018). These batteries shared the same experimental conditions of temperature and discharge current, although their stopping voltages were different. Other studies, such as Tagade et al. [64], Ng et al. [37], and Patil et al. [43], built up predictive models for more than these four batteries, but focused on only the prediction of battery capacity. Sbarufatti et al. [59] considered the prediction with a cycle, that is, to predict the discharge voltages after a time point given the discharge voltage observations prior to the time point within the same discharge cycle. Recently, Nascimento et al. [36] studied 12 out of the 20 batteries and combined recurrent neural networks with physics-based knowledge on battery aging to predict discharge curves for a battery. However, they did not take into account of the different experimental conditions that could affect the aging behaviors of batteries, which is also the common drawback in the other work mentioned above.

There are useful insights we can gain from the work above. For example, Saha and Goebel [55] mentioned the importance of considering the rest period between discharge cycles. This is denoted as Δ_{ic}, which is the elapsed time between cycle c and $c-1$ of battery i. The effect of Δ_{ic} is known as self-recharge, a phenomenon where battery capacity increases after resting. This is due to the dissipation of build-up materials around electrodes. The authors

suggest accounting for Δ_{ic} in the form of an exponential process. Since an increase in battery capacity leads to a larger EOD, we use $\exp(-1/\Delta_{ic})$ as a covariate in modeling the EOD. Saha and Goebel [55], Liu et al. [27], and Liu et al. [30] also considered autoregressive effects in their modeling of battery capacity and found that the inclusion of an order-one autoregressive term in the EOD model would significantly increase its prediction power.

In the next section, we describe the details of our functional degradation models. Instead of relying on historical capacity measurements, we take full advantage of all the information contained in voltage curves from the past discharge cycles to make predictions of the entire voltage discharge curves in future discharge cycles. Predicted VDCs allow battery users the access of specific features that can be extracted from a complete VDC. For example, besides the EOD of a discharge cycle, the decomposition of a VDC curve into components for an impedance model can be useful for discharge performance evaluation and voltage unloading within the cycle [32]. Once voltage discharge curves are available, degradation measure can be calculated based on the chosen health indicator, leading to more versatile battery health prognostics as demonstrated in Section 4.5.

4.3 Functional Degradation Model

4.3.1 Notation

Suppose that there are n batteries. For battery i, n_i cycles are tested with each cycle producing a discharge curve. Let $y_{ic}(r)$ be the original discharge curve for battery i at cycle c, $c = 1, \cdots, n_i$, and $i = 1, \cdots, n$. Here the discharge time $r \in [0, b_{ic}]$, with b_{ic} being the end time for this cycle. The technical term for b_{ic} in the context of battery discharge is end of discharge time (EOD). The experimental conditions are collectively represented by the

covariate vector $\boldsymbol{z}_i$ of length m, which includes the temperature, the discharge current, and the stopping voltage.

To make the discharge curves comparable across different cycles, we first re-scale the discharge time by the end time to obtain a scaled curve. Namely, we consider the re-scaled time $t = r/b_{ic}$ such that $t \in [0, 1]$. And the scaled curve is $x_{ic}(t) = y_{ic}(r) = y_{ic}(b_{ic}t)$. Now each of the original discharge curves is represented by the pair $(b_{ic}, x_{ic}(\cdot))$, with b_{ic} being the EOD. Then the observations for battery i become $(b_{ic}, x_{ic}(\cdot), \boldsymbol{z}_i)$, a mix of a scalar component, a functional component and a vector component. The first three plots in Figure 4.3 provides an illustration of this data processing step.

Our goal is to model $y_{ic}(r)|\boldsymbol{z}_i$, that is, to model $(b_{ic}, x_{ic}(t))|\boldsymbol{z}_i$, a response consisting of the scalar component b_{ic} and the functional component $x_{ic}(t)$. Based on a conditional distribution argument, this is equivalent to modeling $b_{ic}|x_{ic}(t), \boldsymbol{z}_i$ and $x_{ic}(t)|\boldsymbol{z}_i$.

4.3.2 Step 1: Predictive modeling of $x_{ic}(t)$

The first step is to develop a predictive model for the scaled curves $x_{ic}(\cdot)$ with the incorporation of the covariates $\boldsymbol{z}_i$. The model must have the ability to predict the scaled curve $x_{ic}(\cdot)$ at a future cycle c. This is achieved through performing a functional principal component analysis (FPCA) on $x_{ic}(\cdot)$ and then develop a linear regression model with the FPC scores as the responses against the cycle index number c and experimental conditions $\boldsymbol{z}_i$.

Suppose $\Sigma(s, t)$ is the covariance function for the random functions $x_{ic}(t)$. Let λ_j and $\phi_j(t)$, $j = 1, 2, \ldots$, be the eigenvalues and eigenfunctions of $\Sigma(s, t)$. Then $x_{ic}(t)$ admits the Karhunen-Loève expansion

$$x_{ic}(t) = \mu(t) + \sum_{j=1}^{\infty} \gamma_{ic,j}\phi_j(t), \tag{4.1}$$

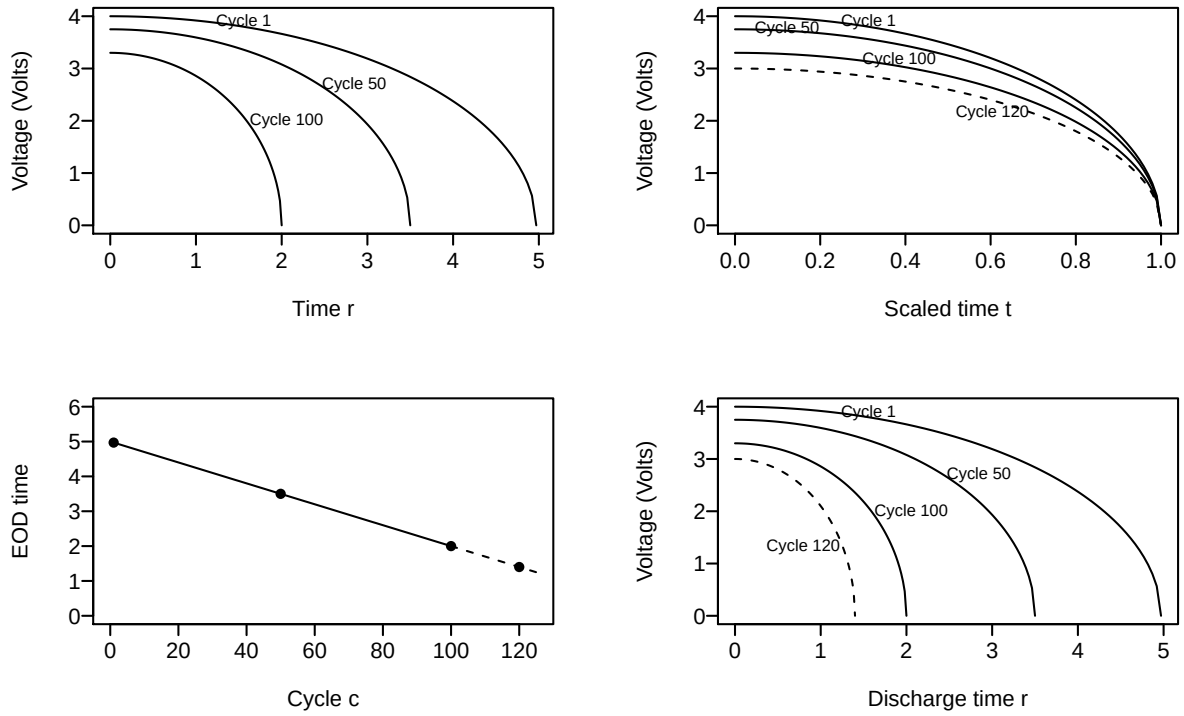


Figure 4.3: Diagram of voltage discharge curve processing for a toy example of battery. Top left: VDCs $y_{ic}(r)$ from three example cycles ($c = 1, 50, 100$) plotted on their original discharge time domain (solid curves). Top right: scaled VDCs $x_{ic}(t)$ with discharge time rescaled by the EODs. Dashed line is the predicted curve for cycle 120. Bottom left: EODs b_{ic} plotted against the cycle index c. Dashed part is the predicted EOD after cycle 100. Bottom right: Predicted VDC for cycle 120 superimposed on the plot of observed VDCs on the discharge time domain.

where $\mu(t)$ is the mean function and $\gamma_{ic,j} = \int_0^1 x_{ic}(t)\phi_j(t)dt$ are uncorrelated random coefficients with mean 0 and variance λ_j. In practice, a truncated expansion is often used such that

$$x_{ic}(t) \approx \mu(t) + \sum_{j=1}^{K} \gamma_{ic,j}\phi_j(t), \tag{4.2}$$

where K is a pre-selected truncation point. Let $\boldsymbol{\gamma}_{ic} = (\gamma_{ic,1}, \ldots, \gamma_{ic,K})^\top$. We consider the

multivariate mixed effects model

$$\boldsymbol{\gamma}_{ic} = \mathbf{v}_0 + \mathbf{u}_{0i} + (\mathbf{v}_1 + \mathbf{u}_{1i})c + \boldsymbol{P}\boldsymbol{z}_i + \boldsymbol{\delta}_{ic}, \tag{4.3}$$

where $\mathbf{v}_0 = (v_{01}, \ldots, v_{0K})^\top$ is the vector of intercepts, $\mathbf{u}_{0i} = (u_{01i}, \ldots, u_{0Ki})^\top$ is the vector of random intercepts for battery i, $\mathbf{v}_1 = (v_{11}, \ldots, v_{1K})^\top$ is the vector of slopes, $\mathbf{u}_{1i} = (u_{11i}, \ldots, u_{01i})^\top$ is the vector of random slopes for battery i, $\boldsymbol{P} = (\mathbf{p}_1, \ldots, \mathbf{p}_m)$ is a K by m matrix with column h being $\mathbf{p}_h = (p_{h1}, \ldots, p_{hK})^\top$, the vector of slopes for covariate h, $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{im})^\top$ is the vector of m covariates for battery i, and $\boldsymbol{\delta}_{ic} = (\delta_{ic,1}, \ldots, \delta_{ic,K})^\top$ is the vector of random errors.

For the implementation of this part, we first perform the FPCA on all the scaled curves $x_{ic}(t)$, take the first K FPCs $\{\hat{\phi}_j(t) : j = 1, \ldots, K\}$, then compute the K FPC scores $\{\hat{\gamma}_{ic,j}, j = 1, \ldots, K\}$ corresponding to each curve $x_{ic}(t)$. These FPC scores are then fed into the multivariate mixed effects model (4.3) as responses. The parameters to be estimated include the fixed-effects components, $\{\boldsymbol{v}_0, \boldsymbol{v}_1, \boldsymbol{P}\}$, and the distributional parameters of the random effects. Suppose $\boldsymbol{U} = (\mathbf{u}_{0i}^\top, \mathbf{u}_{1i}^\top)^\top \sim \mathcal{N}(\mathbf{0}, \Sigma_U)$ and $\boldsymbol{\delta}_{ic} \sim \mathcal{N}(\mathbf{0}, \Sigma_\delta)$. We can apply the standard tool for multivariate linear mixed effects models (MLMMs) to estimate the fixed effects $\{\boldsymbol{v}_0, \boldsymbol{v}_1, \boldsymbol{P}\}$, and the covariance matrices Σ_U and Σ_δ of the random effects. Let $\boldsymbol{\Gamma}_i = (\boldsymbol{\gamma}_{i.,1}, \boldsymbol{\gamma}_{i.,2}, \ldots, \boldsymbol{\gamma}_{i.,K})$, where each $\boldsymbol{\gamma}_{i.,k}$ is the $n_i \times 1$ vector of the k^{th} FPC scores for the n_i cycles of battery i. Following the suggestions in Shah et al. [61] and Thiébaut et al. [66], we vectorize the response matrix into $\boldsymbol{y}_i = \mathbf{vec}(\Gamma_i) = (\boldsymbol{\gamma}_{i.,1}^\top, \boldsymbol{\gamma}_{i.,2}^\top, \ldots, \boldsymbol{\gamma}_{i.,K}^\top)^\top$. Similarly, rewrite $\boldsymbol{\delta}_{ic}$ as $\boldsymbol{e}_i = (\boldsymbol{\delta}_{i.,1}^\top, \boldsymbol{\delta}_{i.,2}^\top, \ldots, \boldsymbol{\delta}_{i.,K}^\top)^\top$ with $\boldsymbol{\delta}_{i.,j} = (\delta_{i1,j}, \ldots, \delta_{in_i,j})^\top$. Then the model becomes

$$\boldsymbol{y}_i = \boldsymbol{X}_i^* \boldsymbol{v} + \boldsymbol{Z}_i^* \boldsymbol{u}_i + \boldsymbol{e}_i. \tag{4.4}$$

Here $\boldsymbol{X}_i^*$ is the $Kn_i \times q^*$ design matrix collecting all the predictors such as cycle number,

experimental covariates $\boldsymbol{z}_i$, and indicator vector representing the membership, among the K components, of the FPC score in the response. Matrix $\boldsymbol{Z}_i^*$ with dimension $Kn_i \times r^*$ is the random design matrix. Vector $\boldsymbol{v}$ is the vectorized version of the fixed effect parameters $\{\boldsymbol{v}_0, \boldsymbol{v}_1, \boldsymbol{P}\}$, while $\boldsymbol{u}_i$ collects individual random effects. We then estimate model (4.4) via the maximum likelihood or restricted maximum likelihood (REML) approach.

4.3.3 Step 2: Modeling $b_{ic}|x_{ic}(t)$ and its parameter estimation

The EOD b_{ic} is a scalar measurement with a longitudinal nature. We consider several versions of mixed effects models here. The initial model is

$$b_{ic} = \alpha_0 + w_{0i} + (\alpha_1 + w_{1i})c + \epsilon_{ic}, \tag{4.5}$$

where α_0 and α_1 are the intercept and slope in the fixed effects, w_{0i} and w_{1i} are the random intercept and slope for battery i, and ϵ_{ic} are random errors with mean 0 and variance σ_ϵ^2. The random effects, $\mathbf{w}_i = (w_{0i}, w_{1i})$ are assumed to follow $N(0, \sigma_\epsilon^2 \boldsymbol{\Psi})$.

Next, we consider two models with the inclusion of covariates into model (4.5). Both of these models contain the experimental conditions $\boldsymbol{z}_i$. In addition, the second model also includes the voltage discharge curve $x_{ic}(\cdot)$ as a functional covariate.

$$b_{ic} = \alpha_0 + w_{0i} + (\alpha_1 + w_{1i})c + \boldsymbol{z}_i^\top \boldsymbol{\alpha}_z + \epsilon_{ic}, \tag{4.6}$$

$$b_{ic} = \alpha_0 + w_{0i} + (\alpha_1 + w_{1i})c + \int_0^1 (\beta(t) + b_i(t))\, x_{ic}(t)dt + \boldsymbol{z}_i^\top \boldsymbol{\alpha}_z + \epsilon_{ic}, \tag{4.7}$$

where $\boldsymbol{\alpha}_z = (\alpha_2, \alpha_3, \dots)^\top$ is the coefficient vector for experimental conditions, and $\beta(t)$ is the unknown coefficient function modeling the fixed-effect of the scaled voltage discharge curve, and $b_i(t)$ is the random coefficient function for battery i, assumed to be a Gaussian process

with mean 0 and covariance function $\gamma(s,t)$. In both (4.6) and (4.7), we assume ϵ_{ic} is i.i.d random variable following a normal distribution with mean 0 and variance σ_ϵ^2. Model (4.6) is a standard linear mixed effects model, and can be estimated through standard software procedures for linear mixed-effects models.

Model (4.7) is a type of functional linear mixed effects (FLMM) model. Liu et al. [25] suggested a expectation-maximization (EM) REML-based algorithm to fit a functional linear model that allows mixed effects for both scalar and functional covariates. To fit this model, we can expand $\beta(t)$ and $b_i(t)$ as linear combinations of basis functions, e.g., B-spline basis functions. Suppose the expansions are respectively $\beta(t) = \sum_{j=1}^{R} m_j\phi_j(t) = \boldsymbol{\phi}^\top(t)\mathbf{m}$ and $b_i(t) = \sum_{k=1}^{S} q_{ik}\psi_k(t) = \boldsymbol{\psi}^\top(t)\mathbf{q}_i$. Since $b_i(t)$ is a zero-mean Gaussian process, we can assume the coefficient vector $\mathbf{q}_i \sim N(0, \sigma_\epsilon^2\mathbf{D})$ for some covariance matrix $\mathbf{D}$ such that $\gamma(s,t) = \sigma_\epsilon^2\boldsymbol{\psi}^\top(s)\mathbf{D}\boldsymbol{\psi}(t)$. Upon substituting the expansions of $\beta(t)$ and $b_i(t)$, We can rewrite (4.7) as

$$b_{ic} = \mathbf{G}_{ic}^\top\boldsymbol{\alpha} + \mathbf{H}_{ic}^\top\mathbf{w}_i + \mathbf{J}_{ic}^\top\mathbf{m} + \mathbf{K}_{ic}^\top\mathbf{q}_i + \epsilon_{ic}, \tag{4.8}$$

where $\mathbf{G}_{ic} = (\mathbf{1}, c, \boldsymbol{z}_i)^\top$, $\mathbf{H}_{ic}$ is the matrix collecting the corresponding random components, $\mathbf{J}_{ic} = \int_0^1 x_{ic}(t)\boldsymbol{\phi}(t)$ and $\mathbf{K}_{ic} = \int_0^1 x_{ic}(t)\boldsymbol{\psi}(t)$. Then the task becomes estimating the basis expansion coefficients $\mathbf{m} = (m_1, \ldots, m_R)^\top$ and $\mathbf{q}_i = (q_{i1}, \ldots, q_{iS})^\top$, the intercept and slopes $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \boldsymbol{\alpha}_z^\top)^\top$ in the fixed effect, and the variance components σ_ϵ^2, $\boldsymbol{\Psi}$ and $\mathbf{D}$. Given σ_ϵ^2, $\boldsymbol{\Psi}$ and $\mathbf{D}$, the objective function for $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \mathbf{m}^\top)^\top$ and $\boldsymbol{\vartheta}_i = (\boldsymbol{w}_i^\top, \mathbf{q}_i^\top)^\top$ is

$$\begin{aligned} M(\boldsymbol{\theta}, \boldsymbol{\vartheta}_i) = &\sum_{i=1}^{n}\sum_{j=1}^{n_i} \frac{1}{2\sigma_\epsilon^2}\left(b_{ic} - \mathbf{G}_{ic}^\top\boldsymbol{\alpha} - \mathbf{H}_{ic}^\top\mathbf{w}_i - \mathbf{J}_{ic}^\top\mathbf{m} - \mathbf{K}_{ic}^\top\mathbf{q}_i\right)^2 \\ &+ \frac{\lambda_\beta}{2}\int_0^1 \{\beta''(t)\}^2 dt + \frac{\lambda_b}{2}\sum_{i=1}^{n}\int_0^1 \{b_i''(t)\}^2 dt \\ &+ \frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\mathbf{q}_i^\top\mathbf{D}^{-1}\mathbf{q}_i + \frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\boldsymbol{w}_i^\top\boldsymbol{\Psi}\boldsymbol{w}_i. \end{aligned} \tag{4.9}$$

This objective function comprises of a least squares component, the roughness penalties for $\beta(t)$ and $b_i(t)$ with λ_β and λ_b be the corresponding smoothing parameters, and the REML components for the random components $\mathbf{w}_i = (w_{0i}, w_{1i})^\top$ and $\mathbf{q}_i$. Note that λ_b actually absorbs a term of σ_ϵ^{-1} which is supposed to appear in the penalty for b_i. Let $\mathbf{b}_i = (b_{i1}, \ldots, b_{in_i})^\top$, $\tilde{\mathbf{G}}_i = (\tilde{\mathbf{G}}_{i1}, \ldots, \tilde{\mathbf{G}}_{in_i})^\top$ with $\tilde{\mathbf{G}}_{ic} = (\mathbf{G}_{ic}^\top, \mathbf{J}_{ic}^\top)^\top$, and $\tilde{\mathbf{H}}_i = (\tilde{\mathbf{H}}_{i1}, \ldots, \tilde{\mathbf{H}}_{in_i})^\top$ with $\tilde{\mathbf{H}}_{ic} = (\mathbf{H}_{ic}^\top, \mathbf{K}_{ic}^\top)^\top$. Define the matrices $G_\phi = \int_0^1 \frac{d^2\phi(t)}{dt^2}\left(\frac{d^2\phi(t)}{dt^2}\right)^\top dt$ and $G_\psi = \int_0^1 \frac{d^2\psi(t)}{dt^2}\left(\frac{d^2\psi(t)}{dt^2}\right)^\top dt$. Then a full matrix-vector form of the objective function (4.9) is

$$\begin{aligned} M(\boldsymbol{\theta}, \boldsymbol{\vartheta}_i) = &\sum_{i=1}^{n}\sum_{j=1}^{n_i} \frac{1}{2\sigma_\epsilon^2} \parallel \mathbf{b}_i - \tilde{\mathbf{G}}_i\boldsymbol{\alpha} - \tilde{\mathbf{H}}_i\boldsymbol{\vartheta}_i \parallel^2 + \frac{\lambda_\beta}{2}\mathbf{m}^\top G_\phi \mathbf{m} + \frac{\lambda_b}{2}\sum_{i=1}^{n}\int_0^1 \mathbf{q}_i^\top G_\psi \mathbf{q}_i \\ &+ \frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\mathbf{q}_i^\top \mathbf{D}^{-1}\mathbf{q}_i + \frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\boldsymbol{w}_i^\top \boldsymbol{\Psi}^{-1}\boldsymbol{w}_i \end{aligned} \tag{4.10}$$

Objective function (4.10) can be minimized with a closed-form solutions for $\boldsymbol{\alpha}$ and $\boldsymbol{\vartheta}_i$. To estimate σ_ϵ^2 and $\mathbf{D}$, we can employ an iterative EM REML-based starting with some initial values for $\sigma_\epsilon^{2(0)}$ and $\mathbf{D}^{(0)}$. Then the M-step computes the estimates of $\boldsymbol{\alpha}^{(r)}$ and $\boldsymbol{\vartheta}_i^{(r)}$ as the minimizer of (4.10), followed by the E-step of updating $\sigma_\epsilon^{2(r)}$ and $\mathbf{D}^{(r)}$. The procedure is repeated until all estimators converge within a pre-specified tolerance level. We employ multi-fold cross validation for both the selection of smoothing parameters λ_β and λ_b and the selection of models and covariates to include in $\boldsymbol{z}_i$.

4.3.4 Step 3: Estimation, prediction, and degradation analysis of $y_{ic}(r)$

Once the scaled discharge curve $x_{ic}(t), t \in [0, 1]$, and the EOD b_{ic} are estimated through the two steps described above, the estimate for the standard discharge curve $y_{ic}(r)$ on its natural domain, $r \in [0, b_{ic}]$, can be constructed point-wisely by simply re-scaling the estimate $\hat{x}_{ic}(t)$

with the EOD estimate $\hat{b}_{ic}$. The prediction of the voltage discharge curve $\hat{y}_{i\tilde{c}}(r)$ for a future cycle $\tilde{c}$ can be obtained similarly through the predictions $\hat{x}_{i\tilde{c}}(t)$ and $b_{i\tilde{c}}$. Then $\hat{y}_{i\tilde{c}}(r)$ is used as a functional measurement prediction for the following degradation analysis.

One advantage of producing predicted VDCs in the fully functional form is the flexibility in performing degradation analysis according to the degradation definition set by practitioners. One can examine the full predicted VDCs and anticipate discharge voltage behavior within a future discharge cycle. We illustrate this approach in more details in Section 4.5.3. Another popular approach is to convert a full VDC into a scalar degradation measurement. As demonstrated in Section 4.4, existing methods construct degradation path models directly from a scalar degradation measurement Υ_{ic} calculated from the observed VDC, $y_{ic}(r)$. Such scalar measurement could be the L^p-norm of the $y_{ic}(r)$

$$\Upsilon_{ic} = ||y_{ic}(r)||_p \equiv \left(\int_0^{b_{ic}} |y_{ic}(r)|^p dr \right)^{1/p} \tag{4.11}$$

The power p is pre-specified. When $p = 1$, Υ_{ic} is the area under the discharge curve, which is directly related to battery capacity. To compare a battery's degradation at cycle c relative to its original state, i.e., the first discharge cycle, we can further compute the degradation amount

$$d_{ic} = \frac{\Upsilon_{ic} - \Upsilon_{i1}}{\Upsilon_{i1}} = \frac{||y_{i1}(r)||_p - ||y_{ic}(r)||_p}{||y_{i1}(r)||_p} \tag{4.12}$$

The degradation amount d_{ic} in (4.12) can be interpreted as the difference between the L^p-norm of the discharge curve at cycle c and the L^p-norm of the discharge curve at the first cycle for battery i. This measure can be compared to any soft failure threshold D_f set by practitioners to represent the end of rechargeable battery usability. Figure 4.4 demonstrate the degradation analysis procedure on a toy example. We consider three hypothetical batteries or units (Panels 1 to 3). Each unit starts with some observed VDCs (solid curves), which

yield a predicted VDC (dashed curve) based on a prediction model. We then computes a degradation amount for each VDC (observed or predicted). Then the degradation amounts for each battery are assembled to produce the degradation path in Panel 4. At last, the degradation path can be compared with the soft failure threshold to determine the status of the unit.

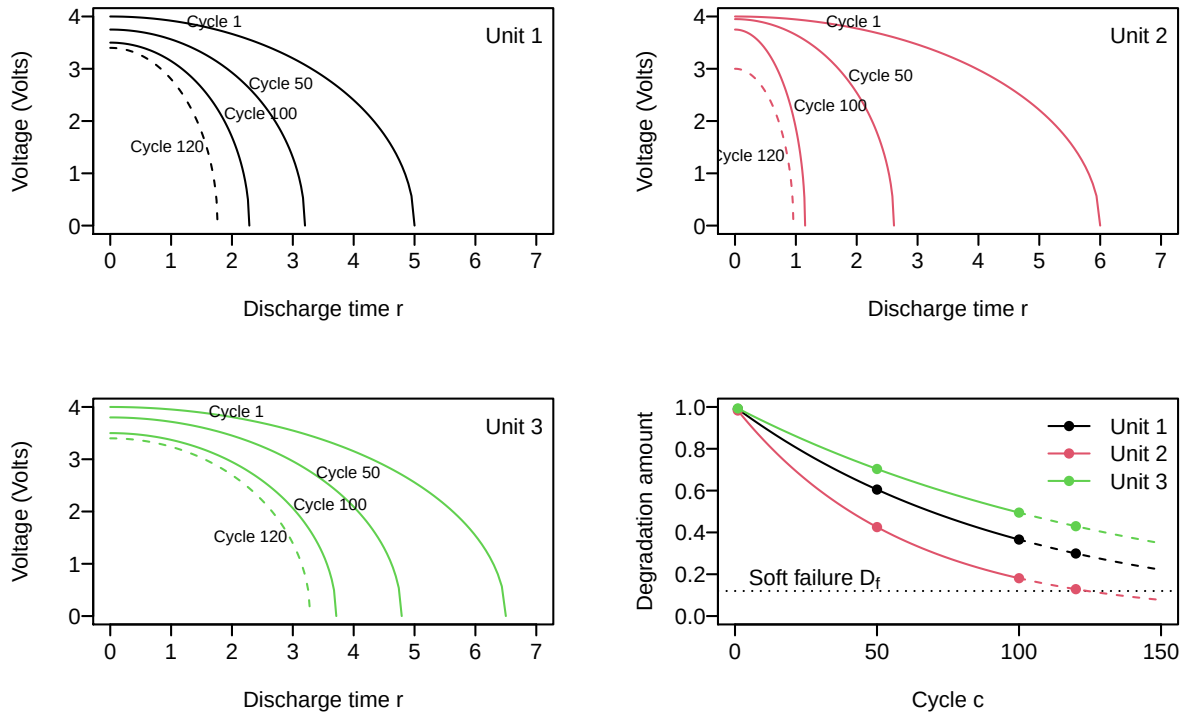


Figure 4.4: Diagram for degradation analysis for multiple batteries simultaneously. Panel 1 (top left), Panel 2 (top right), and Panel 3 (bottom left): Voltage discharge curves of each unit (battery). Solid curves are for modeling under procedure proposed in 4.3. Dashed curves are predictions based on final model. For each of these VDCs, degradation amount is measured according to equation (4.12). Panel 4 (bottom right): degradation paths for all units obtained from VDCs with observed (solid) and predicted (dashed) values overlayed with a soft failure threshold.

4.4 Simulation Study

4.4.1 Existing degradation models

Current methods in degradation analysis of rechargeable batteries do not take into account the full functional form of VDCs. The observed VDCs are transformed into a scalar degradation measurement through equations (4.11) and (4.12). Then a general path model [34] is constructed based on the observed degradation amount d_{ic}

$$d_{ic} = \mathcal{D}_i(c; \boldsymbol{\beta}, \boldsymbol{\xi}_i | \boldsymbol{z}_i) + \epsilon_{ic}, \tag{4.13}$$

where $\boldsymbol{z}_i$ is the vector of covariates for battery i, $\mathcal{D}_i$ is a function of the cycle c given $\boldsymbol{z}_i$ with parameters $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$. In particular, $\boldsymbol{\beta}$ contains all the fixed parameters common across all batteries and $\xi_i = (\xi_{1i}, \ldots, \xi_{i\ell})^\top$ collects ℓ random parameters representing battery-to-battery variations. In model (4.13), ξ_i is assumed to follow a multivariate normal distribution with mean $\mathbf{0}$ and covariance $\boldsymbol{\Sigma}$, and the random errors ϵ_{ic} are assumed to be independent and identically distributed, following a normal distribution with mean 0 and variance σ_ϵ^2.

The first step in the degradation model (4.13) is specifying the functional form of $\mathcal{D}_i(c)$ based on the shape of the degradation path formed by the degradation amounts d_{ic} along the cycles. Based on Figure 4.2, we assume a linear degradation path of the form

$$\mathcal{D}_i(c) = \beta_0 + (\beta_1 + \xi_{i1})c + \boldsymbol{z}_i^\top \beta_z. \tag{4.14}$$

By setting $\exp(\zeta_{i1}) = |1 + \xi_{i1}/\beta_1|$, we can see that model (4.14) has an equivalent form

$$\mathcal{D}_i(c) = \beta_0 + \beta_1 \exp(\zeta_{i1})c + \boldsymbol{z}_i^\top \beta_z, \tag{4.15}$$

where β_1 may differ from β_1 in (4.14) by a sign and the random effect ζ_{i1} now assumes a lognormal probability distribution. The goal is now to estimate parameters $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \sigma_\epsilon^2, \boldsymbol{\Sigma}\}$. Note that in general model (4.13) is a nonlinear mixed effects model, which can be estimated through standard methods such as ML or REML. In practice, we apply the `nlme` and `lme` functions from the R package `nlme` [44].

4.4.2 Simulation settings

We conduct a numerical simulation study to compare degradation analysis performance between our functional data approach and existing methods. In numerical studies, we simulate data from a hypothetical experiment with n number of units. For each $i = 1, \ldots, n$, unit i is set to go through n_i cycles of action that result in its degradation. Within each cycle, a curve observation is simulated. This curve is assembled from two components: a curve on the standard $[0, 1]$ domain and a domain end point that scales the standard curve to an individualized domain. In particular, the domain end point b_{ic} for curve c of unit i is simulated from the following model

$$b_{ic} = \alpha_0 + w_{0i} + (\alpha_1 + w_{1i})c + \alpha_2 x_i + \alpha_3 b_{i,c-1} + \alpha_4 \exp\left\{\frac{-1}{\Delta_{ic}}\right\} + \epsilon_{ic}, \tag{4.16}$$

where $w_{0i} \sim \mathcal{N}(0, \sigma_0^2)$, $w_{1i} \sim \mathcal{N}(0, \sigma_1^2)$,$\epsilon_{ic} \sim \mathcal{N}(0, \sigma_\epsilon^2)$. The covariate x_i represents a testing condition for unit i, which is generated from a uniform distribution on $[0, 1]$. The time lag Δ_{ic} between cycles $c-1$ and c is set up as follows. When c is divisible by 20, we set $\Delta_{ic} = 100+\ell$ where $\ell \sim \mathcal{N}(0, 2^2)$. Otherwise, $\Delta_{ic} = 5 + \ell$ where $\ell \sim \mathcal{N}(0, 1^2)$. In this manner, we assume the hypothetical system has a long break every 20 cycles. Once b_{ic} is generated, the curve $y_{ic}(r)$ with $r \in [0, b_{ic}]$ is obtained through $y_{ic}(r) = \sum_j^3 \gamma_{ic,j}\phi_{ijc}(r)$, where $\phi_{ijc}(r)$ are the first three basis functions in the cubic B-spline basis system with 5 equally spaced knots on the

interval $[0, b_{ic}]$. Models for the coefficients $\gamma_{ic,j}, j = 1, 2, 3$, are

$$\gamma_{ic,j} = (u_{01} + v_{0ij}) + (u_{11} + v_{1ij})c + u_2 x_i + \delta_{icj} \tag{4.17}$$

where $v_{0ij}, v_{1ij} \sim \mathcal{N}(0, \sigma_v^2)$ and $\delta_{icj} \sim \mathcal{N}(0, \sigma_\delta^2)$.

With the guidance of the application, we choose our parameter values as follows. The fixed-effect coefficients in (4.16) are $\alpha_0 = 2500$, $\alpha_1 = -5$, $\alpha_2 = 50$, $\alpha_3 = 0.5$, and $\alpha_4 = 150$. The variance of random effects and noise are $\sigma_{w0}^2 = 200^2$, $\sigma_{w1}^2 = 2^2$ and $\sigma_\epsilon^2 = 10^2$. In (4.17), the fixed effects are set as $(u_{01}, u_{02}, u_{03}) = (10, 6, 5)$, $(u_{11}, u_{12}, u_{13}) = (-0.03, -0.02, -0.01)$, and $(u_{21}, u_{22}, u_{23}) = (5, 4, 3)$. The variances of random effects and noise are respectively $\sigma_v^2 = 0.01$ and $\sigma_\delta^2 = 1^2$. Figure 4.5 displays data for a unit generated in this manner.

To test the efficacy of the proposed method on the simulated data, we build predictive models from $tr\%$ of the first number of cycles of each unit. Predictions are done on the remaining $1 - tr\%$. We choose the degradation measure of interest to be the normalized difference between the L^1-norms, or areas under the curves, of the VDCs at cycle c and the first cycle, namely, $d_{ic} = \frac{\|y_{i1}(r)\|_1 - \|y_{ic}(r)\|_1}{\|y_{i1}(r)\|_1}$. The accuracy in prediction is evaluated by the root mean squared prediction error RMSPE $= \sqrt{\sum_{i=1}^{n} \sum_{c=\tilde{n}_i}^{n_i} (\hat{d}_{ic} - d_{ic})^2}$ where $\tilde{n}_i = n_i \cdot tr\% + 1$. We also evaluate the goodness of fit in estimating degradation amount using root mean squared error RMSE $= \sqrt{\sum_{i=1}^{n} \sum_{c=1}^{\tilde{n}_i - 1} (\hat{d}_{ic} - d_{ic})^2}$.

Our simulation settings comprise of all the possible combinations of $n = \{20, 50, 100\}$, $n_i = \{50, 100, 150\}$, and train ratio $tr\% = \{50\%, 80\%\}$. Each simulation setting contains 500 data replications. Model for general path modeling is fitted with the `lme` routine from the R package `nlme`. For the simulation study, we apply the linear degradation path in equation (4.14). For the proposed functional model approach, we apply the functional principal component decomposition from `fdapace` R package [5], then the multivariate linear

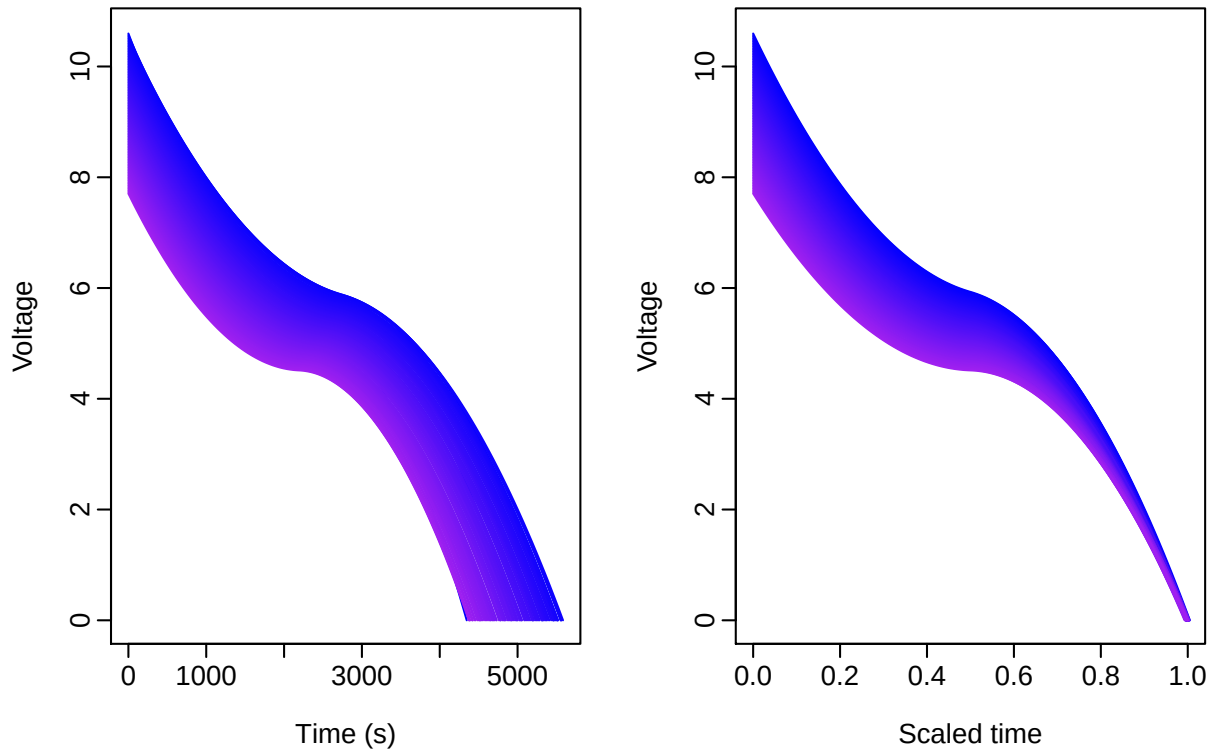


Figure 4.5: Example of simulated discharge curves through 100 discharge cycles for one unit. Colors ranging from blue to purple indicate increase in discharge cycle number. Left panel displays the curves in their original domain. Right panel is the scaled curves.

mixed effects model for FPC scores are constructed using `lme` on the vectorised form as demonstrated above. Two models of estimating EOD time are considered. The first is the linear mixed effects model similar to (4.16). The second is the FLMM where we fit (4.16) with the covariate term $\alpha_2 x_i$ replaced with the functional covariate $\int_0^1 (\beta(t) + b_i(t))\, x_{ic}(t)dt$. We customized the code provided in Liu et al. [25] to obtain the estimated parameters in the functional linear mixed model in model (4.7).

4.4.3 Simulation results

Figure 4.6 shows the boxplots of RMSPEs of the three methods grouped by simulation settings. GPM stands for the general path model using the linear degradation path. Models using functional data approach are pre-fixed with FDM followed by the method of modeling EOD. From Figure 4.6, when training ratio is 80% and number of cycles per battery at 50 and 100, the two FDM models, FDM-LME and FDM-FLMM, produce almost identical results and both are much better than the GPM method.

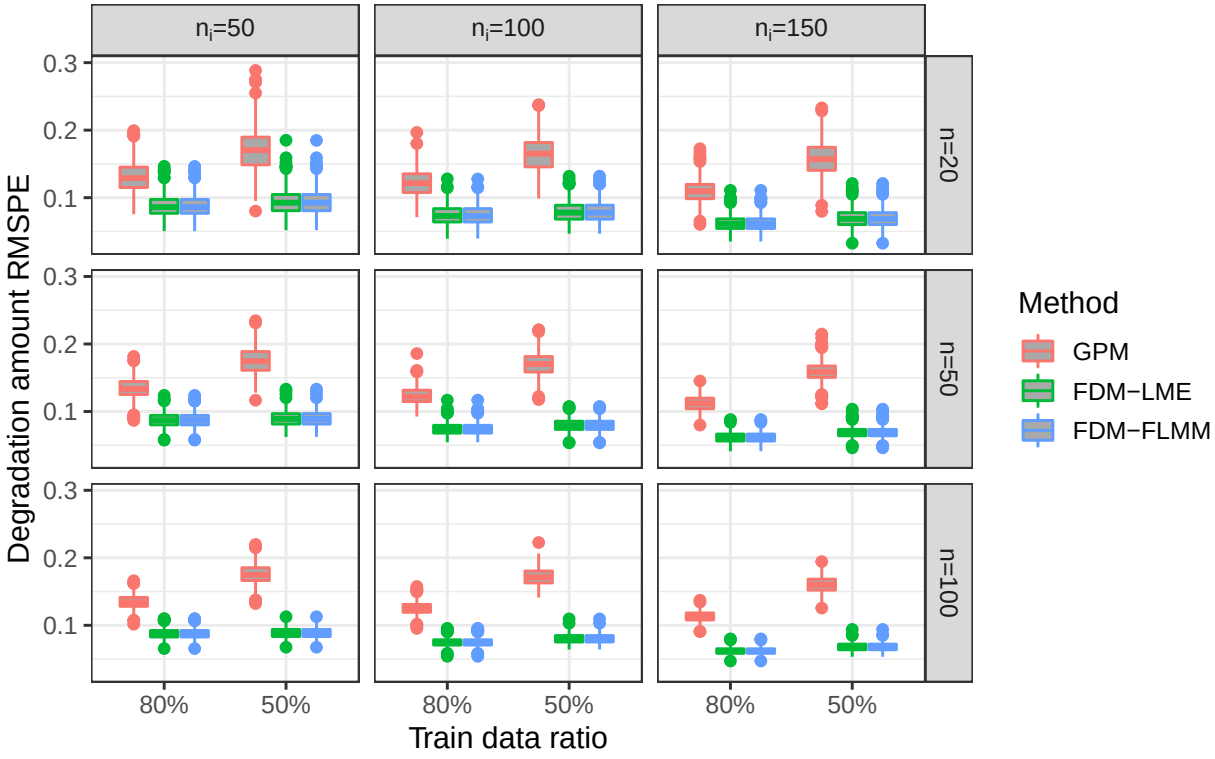


Figure 4.6: RMSPE over all simulation settings combinations from three degradation models.

We examine the goodness of fit for these models on the training data in Figure 4.7. It is not surprising that GPM has a better performance in fitting the training data compared to the two FDM models. The GPM directly models the degradation amounts. Although this yields better fit on training data, it also tends to overfit the data and thus leads to worse

prediction performance as demonstrated above. On the other hand, the two FDM models deal with functional VDCs from two sources: the scaled VDCs and the EODs. Although the FDM approaches do not fit the training data as the direct GPM approach, they gain in the prediction performance which is more important in degradation analysis.

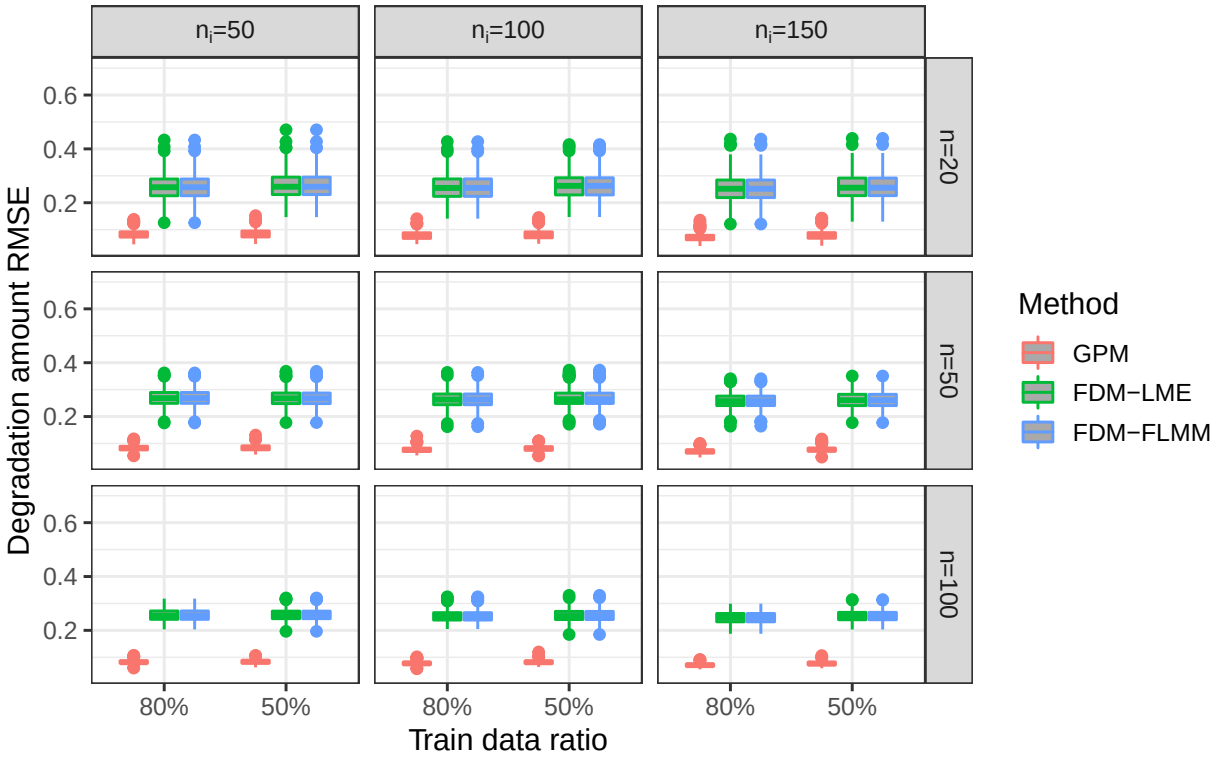
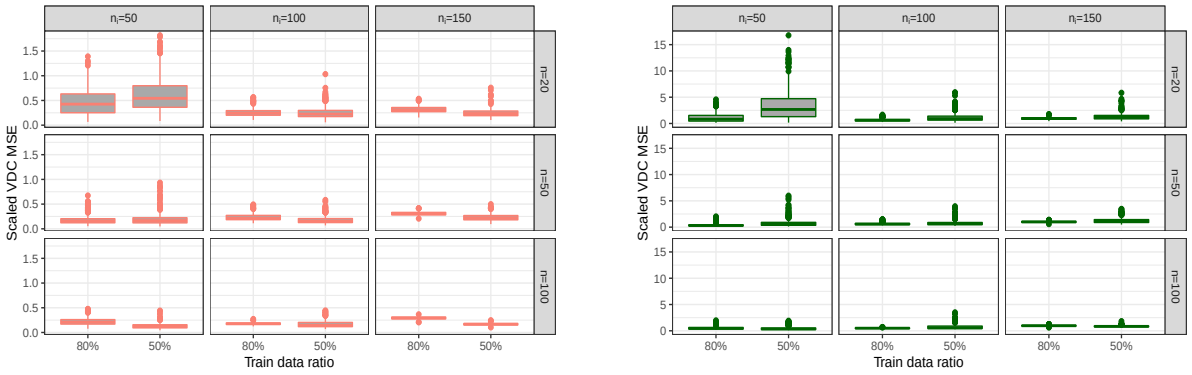


Figure 4.7: Boxplots of RMSE on observed data over all simulation settings combinations from three degradation models.

We examine the functional data models further by looking at the fit and prediction performance for scaled VDCs across different simulation settings in Figure 4.8. Using FPC scores to model scaled VDCs gives good fit based on pointwise MSE over all simulation settings. The disprepancy between fit and prediction MSE reduces as the number n_i of cycles per battery increases. On the other hand, the increase in the number n of batteries leads to smaller variances in prediction errors.

Both FDM-LME and FDM-FLMM use the same method for modeling scaled VDCs through



(a) Training set (b) Testing set

Figure 4.8: Boxplots of Point-wise MSE of scaled VDCs in training (Fit Type) and testing set (Pred. Type) over all simulation settings

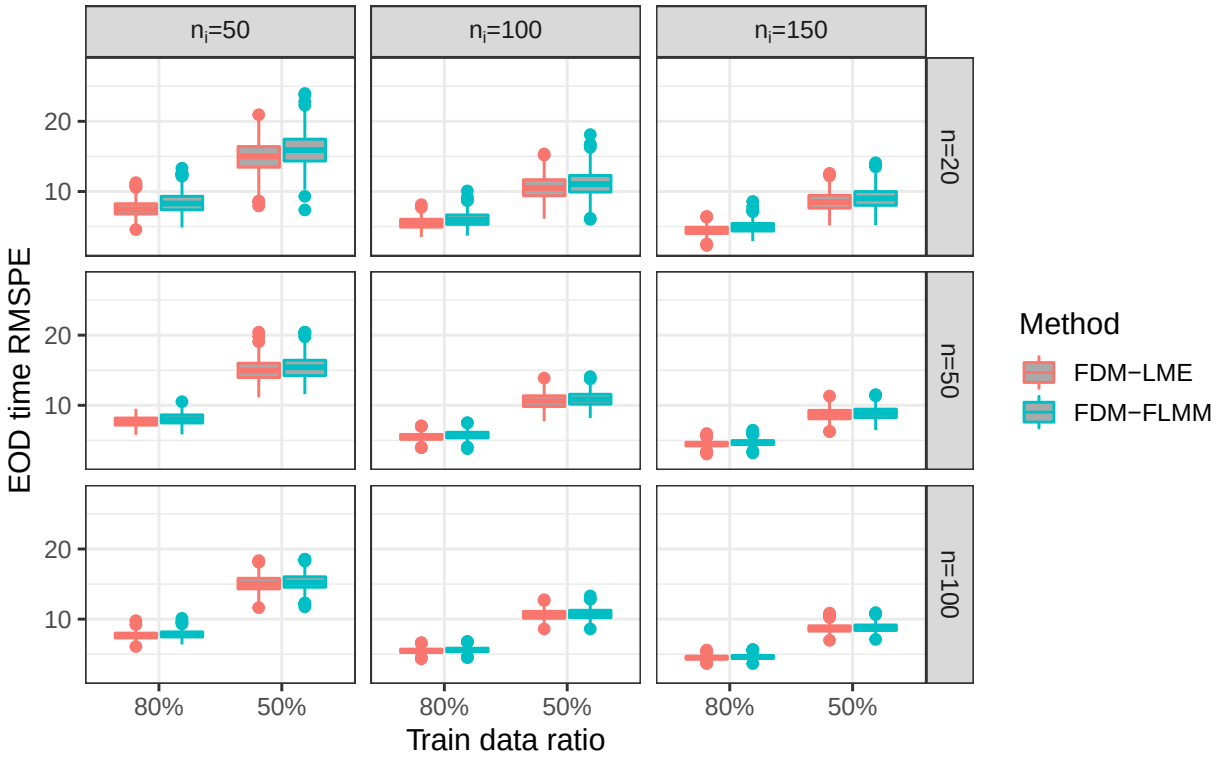


Figure 4.9: Boxplots of RMSPE of EOD time by FDM-LME and FDM-FLMM over all simulation settings.

FPC scores. They differ in modeling the EOD time. Figure 4.9 displays boxplots of RMSPE for predicting EOD time by the two models. While the prediction performance of FDM-

LME is slightly better for the case of small n and n_i, the two methods produce almost identical results as the number of batteries and the number of cycles per battery increase. This explains the similarity in performance between the two models in Figure 4.6.

4.5 Data Analysis

4.5.1 Model construction

The dataset contains voltage discharge curves for $n = 20$ batteries under various experimental conditions. The experimental conditions are respectively the testing temperature (*temp*) in Celsius degrees, discharge current (dc) in Amp, and the stopping voltage (sv) in volts. The number of cycles n_i for a battery ranges from 60 to 200. For each discharge cycle, the end of discharge (EOD) b_{ic} is defined to be the time elapsed from the beginning to the end of the cycle. To obtain the scaled curves $x_{ic}(t)$, we scale the timestamps on each voltage discharge curve (VDC) $y_{ic}(r)$ by the corresponding b_{ic}. Since the time grid for the scaled curve $x_{ic}(t)$ may differ between cycles and batteries, we interpolate each curve to obtain observations on a uniform grid. All the covariates are modeled as continuous variables. We apply the Arrhenius transformation for temperature using room temperature as the baseline level [34], that is

$$z_1 = \frac{11604.52}{temp^{\circ}\mathrm{C} + 273.15} - \frac{11604.52}{24^{\circ}\mathrm{C} + 273.15}$$

For discharge current (dc) and stopping voltage sv, we use the following transformations,

$$z_2 = \frac{\log(dc)}{2} \quad \text{and} \quad z_3 = \frac{\log(sv)}{2}$$

This is called the power law transformation using 2 Amps as baseline level for discharge current and 2 Volts as baseline level for stopping voltage [34]. These are standard transformations in battery studies and can enhance the numerical stability in model fitting. Furthermore, the baseline levels for the three experimental conditions have a covariate value of 0 under these transformations. This makes it easier to interpret the intercept of the model and compare the effects on the response between different levels of each covariate.

When performing the FPCA on the scaled VDCs, we find that the first three FPCs can already explain more than 99% of the total variation in the curves. Therefore, we use $K = 3$ here and represent each curve $x_{ic}(t)$ by the three corresponding FPC scores, $\gamma_{ic,j}$, $j = 1, 2, 3$. We fit the multivariate linear mixed effects model in (4.3) with $\boldsymbol{z}_i = (z_{1i}, z_{2i}, z_{3i})^\top$, using the procedure introduced in Section 4.3.

Once the model is fitted and the parameters are estimated, we apply

$$\hat{\boldsymbol{\gamma}}_{ic^*} = \hat{\boldsymbol{v}}_0 + \hat{\boldsymbol{u}}_{0i} + (\hat{\boldsymbol{v}}_1 + \hat{\boldsymbol{u}}_{1i})c^* + \hat{\boldsymbol{P}}\boldsymbol{z}_i \tag{4.18}$$

to obtain the predicted FPC scores of the scaled voltage discharge curves $x_{ic^*}(t)$ for a future cycle c^*. Combining these scores with the FPCs and the estimated mean discharge function $\hat{\mu}(t)$ yields the prediction of the scaled curve $x_{ic^*}(t)$ for cycle c^*.

The next step is to model the EOD time b_{ic}. Visual inspection in Figure 4.10 indicates that a linear trend in cycle number can capture well the trajectory of EOD time. To capture the variations between batteries, a reasonable model choice is the linear mixed effects model with random terms for the intercept and slope for cycle number grouped by battery. Two additional covariates are introduced following the suggestion in Saha and Goebel [55]. One is the EOD $b_{i,c-1}$ of the previous cycle, to capture the autoregressive nature of EODs found in their analysis. The other is a transformation of the waiting time Δ_{ic} between two cycles,

namely, $\exp(-1/\Delta_{ic})$. Hence the first option to model b_{ic} is the form of model (4.6) with $\boldsymbol{z}_i = (b_{i,c-1}, \exp(-1/\Delta_{ic}), z_{1i}, z_{2i}, z_{3i})^\top$.

A second option is to take advantage of the information from scaled voltage discharge curve by fitting the functional linear mixed effects model (4.7). Since the experimental conditions have already been incorporated into the model for $x_{ic}(t)$, we choose not to include them again in the models for the EOD time, and thus $\boldsymbol{z}_i$ in (4.7) becomes $\boldsymbol{z}_i = (b_{i,c-1}, \exp(-1/\Delta_{ic}))^\top$. We first fit models (4.6) to (4.7) using the training data, that is, the first $tr\%$ of data for each battery. Based on the fitted models, we make predictions on the $1 - tr\%$ held-out testing data.

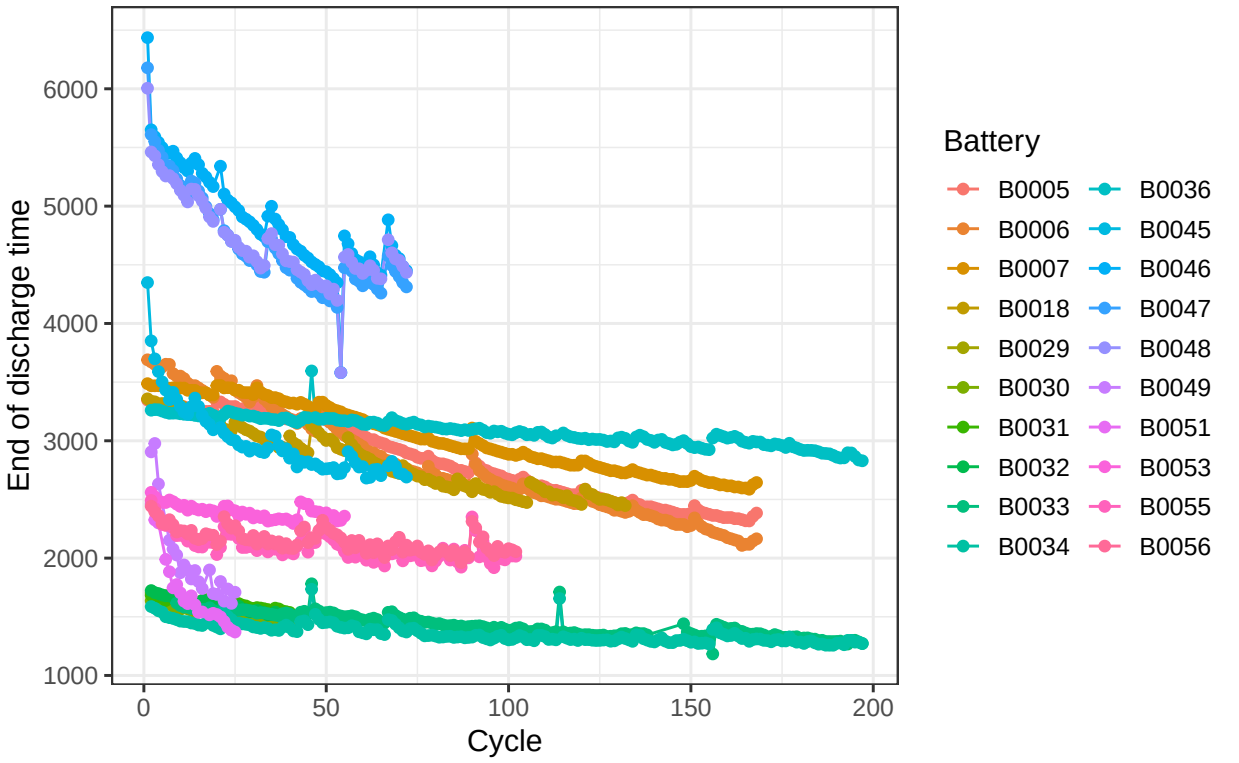


Figure 4.10: End of discharge (EOD) time per cycle by each battery.

After prediction of the scaled curve $x_{ic}(t)$ as well as domain end time b_{ic}, we reconstruct the curve on its natural domain $y_{ic}(s)$ and compute the degradation amount (4.12) using the L^p

norm with $p = 1$. As done in simulations, we compare our degradation predictions to those from the general path model which is based only on the degradation amounts at the training cycles.

4.5.2 Model evaluation

We consider three levels for the train ratio, $tr\% \in \{50\%, 75\%, 90\%\}$. For illustrating purpose, we present the fitted and predicted degradation paths for the 75% train ratio setting. Figure 4.11 displays the predicted degradation paths for all the 20 batteries when 75% of available data for each battery are used in modeling. Predictions from FDM-LME and FDM-FLMM are able to capture well the testing part trajectories of the degradation paths for most batteries while GPM tends to overestimate degradation amounts. For a few exceptions, such as Batteries B0055 and B0056, it is hard to tell which method predicts better. These batteries all have rough degradation paths. While GPM does better prediction for these batteries in the initial period, its prediction deviates more than the FDM methods in the later period. The FDM predictions for these batteries tend to stay in the middle of the oscillating true degradation paths.

Figure 4.12 shows boxplots of absolute errors of the three methods when fitting and predicting degradation amount under the three train/test ratios. In contrast to the under-performance in fitting of the two FDM models versus the GPM in the simulation study, the fits on the training data from all three methods are quite consistent with one another. In prediction, both FDM models give better and more consistent performance than the GPM.

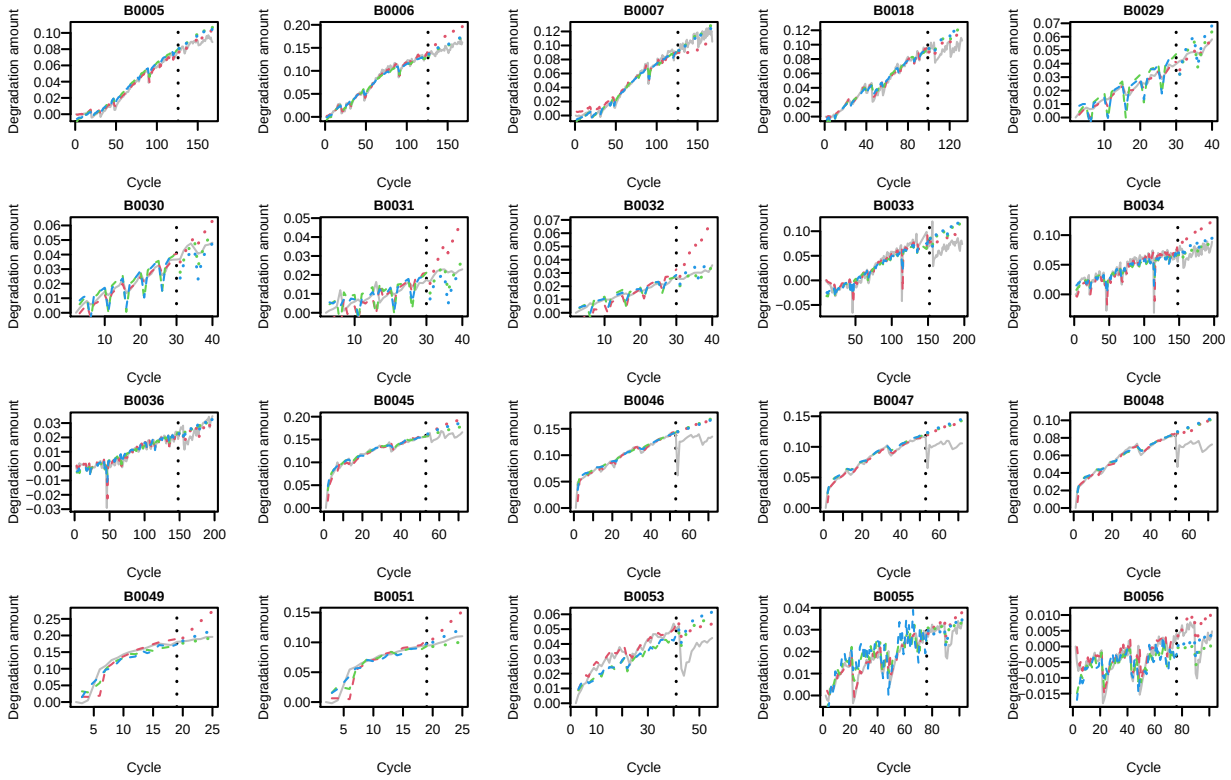


Figure 4.11: Degradation paths for 20 batteries (gray) overlayed with fitted and predicted paths by GPM (red), FDM-LME (green), and FDM-FLMM (blue). Dotted vertical line indicate the separation between train (first 75% of discharge cycles from each battery) and test set (last 25% of discharge cycles from each battery).

4.5.3 Degradation prediction

Lastly, we perform functional degradation analysis using all the available data on the batteries, that is, we make prediction about the battery degradation beyond the discharge cycles available in the data. We use the results from the FDM-FLMM model as an example and predict the VDCs and degradation paths for the next 20 cycles of each battery assuming a 5-hour lag between cycles. Figure 4.13 visualizes the results for two batteries. This routine illustrates the benefits of functional degradation model in producing both the predictions for VDCs and the degradation amount of interest.

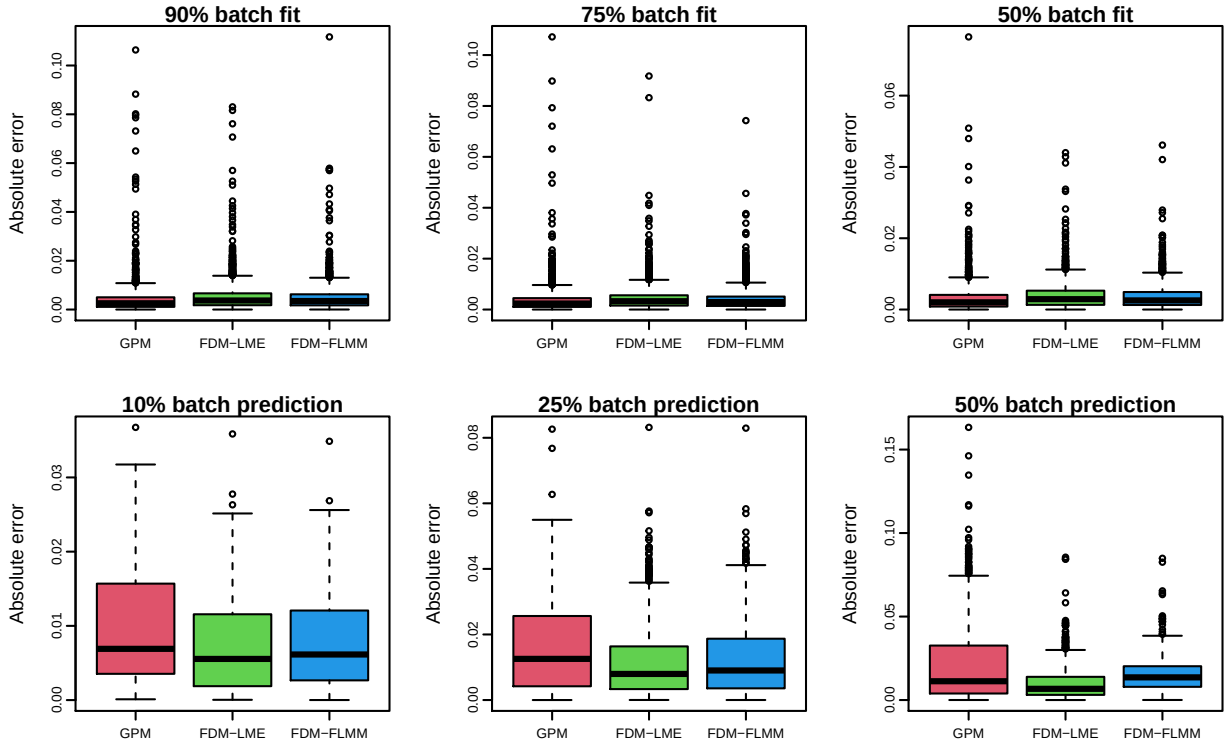


Figure 4.12: Boxplots of absolute error in degradation amount fit and prediction among three considered methods across three train/test split ratios.

4.6 Conclusions and Future Work

In this work, we propose a functional degradation model to analyze a battery life testing study, where voltage discharge curves form a longitudinal set of functional data observations showing a pattern of degradation. Instead of applying standard degradation models directly to degradation measures summarizing functional observations, our functional data approach take full advantage of the whole VDCs. Considering the heterogeneity in the supporting domains of the VDCs, we first re-scale each VDC to reside on the common domain $[0, 1]$ and then simultaneously model the scaled curves and the EOD times through mixed effects models. Simulations show that our functional degradation model can indeed yield better predictions than the standard general path model. Our detailed analysis of the battery life

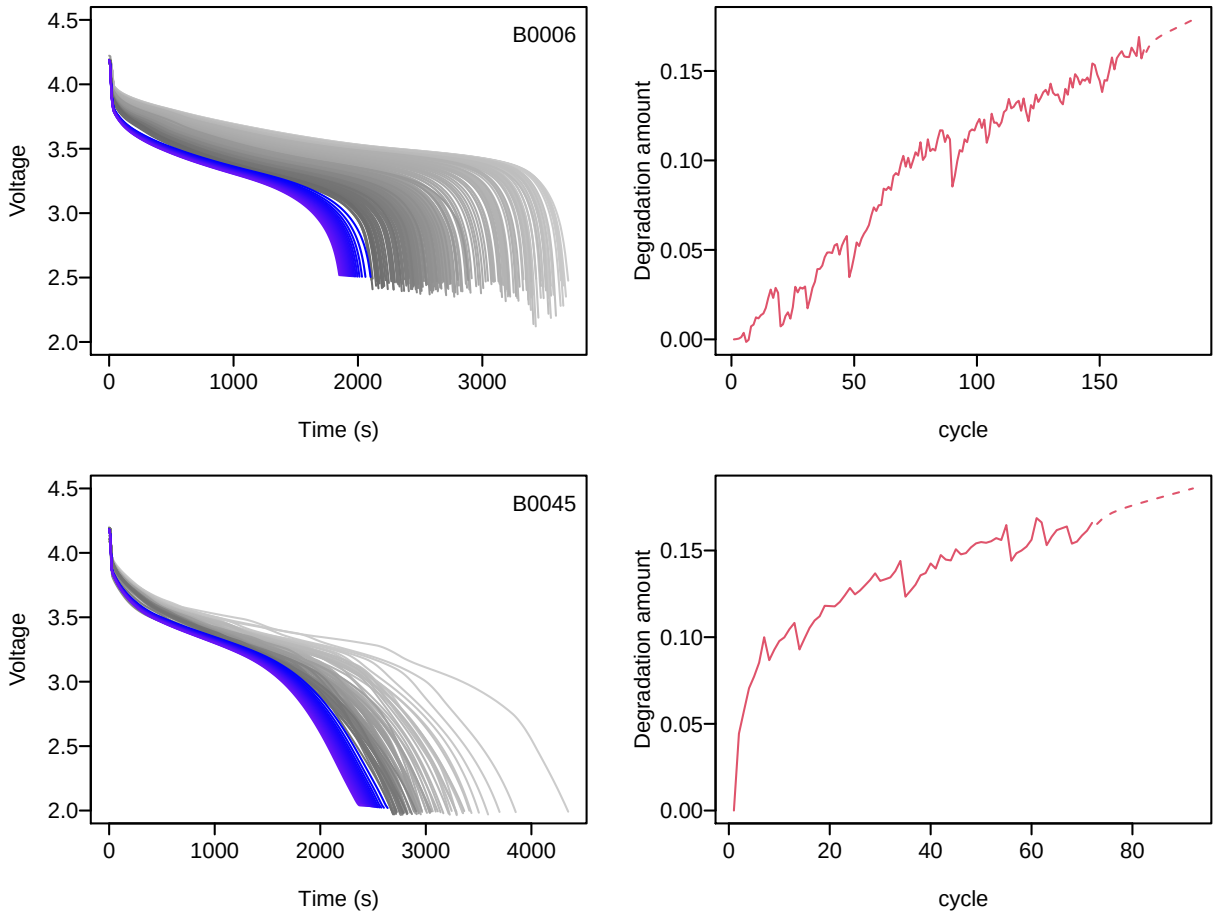


Figure 4.13: Predicted voltage discharge curves and degradation paths for two batteries in the next 20 discharge cycles based on FDM-FLMM model built on entire battery dataset

testing data further reveal the great potential of the functional degradation model for more accurate and more stable degradation prediction for Li-ion batteries.

Another advantage of the functional approach is that its prediction is not only just a predicted value for the degradation amount but also a complete VDC. The availability of complete VDC predictions would allow the practitioners the flexibility of defining their own degradation measures based on the curves. In the future, it will be interesting to consider time-varying covariates such as varying ambient temperature, and multiple loading values for discharge current.

Chapter 5

Conclusions and Future Work

As technology advances, more and more sophisticated data are collected allowing applications and systems to be studied more indept. Functional data analysis is conceived as a field of statistics to deal with data from complex process collected under high frequency. The work in this dissertation contributes to the vastly expanding methodologies of FDA through three main projects. Besides application orientation, an underlying theme of this dissertation is also the study of methods involving functional data observations that are collected longitudinally.

Chapter 2 concentrates on the functional Analysis of Variance (FANOVA) model to compare mean functions among groups of functional data. The focus is on the post-hoc situation where linear contrasts formed betwen group means are put to test. Existing tests focus on situations with independent observations among group members. We propose a new test for when dependency exists amog within-group observations. The test is robust in situation of uncommon covariance functions across groups. Simulation studies prove the that the empirical power of the test is comparable in situation favorable to existing tests. When dependency exists within each group, the proposed test produces stronger power. Drawbacks of the proposed test lie in the size control when it tends to give high type I error rate in small size situation. Additional simulations have shown that this problem is fixed when the number of sample size increases. In the future, it would be beneficial to investigate the performance of the test when within-group dependency is not homogeneous. Furthermore,

so far we only discuss contrast testing for one-way FANOVA model. It will be interesting to generalize the techniques into two-way FANOVA or general designs such as random complete block design.

Chapter 3 follows up from the linear contrast test in FANOVA model with the estimation of the contrast function with local sparsity. This method is especially beneficial for situation where it is of interest to pinpoint regions where contrast function is zero while estimation for non-zero regions maintain accuracy. A direct application is with Raman spectral data. Wavelength positions along Raman spectra indicate the intensity of a specific chemical composition in the liquid sample. The proposed method produce estimated linear contrast constructed from groups of Raman spectral data with wavelength regions of null contrast marking chemical composition of zero influence on contrast.

Chapter 4 explores the power of functional data methodology in the context of degradation analysis. The project is specifically calibrated towards a study on degradation of rechargable batteries under continous charge and discharge. Fully functional approach is taken by coupling two flexible functional linear models for functional response and scalar response with functional covariate. The proposed method is able to produce prediction for future voltage discharge curves. We demonstrate in simulation study the superiority of a functional approach in predicting pre-specified degradation measure compared with standard general path model that work directly scalar measures. We hope this work motivates more application and development of FDA methods in reliability and degradation analysis. A direct expansion would be to examine impact of time-varying covariates toward the degradation of data collected in functional form.

Bibliography

[1] Germán Aneiros, Ricardo Cao, Ricardo Fraiman, Christian Genest, and Philippe Vieu. Recent advances in functional data analysis and high-dimensional statistics. *Journal of Multivariate Analysis*, 170:3–9, 2019. ISSN 10957243. doi: 10.1016/j.jmva.2018.11.007. URL https://doi.org/10.1016/j.jmva.2018.11.007.

[2] Anestis Antoniadis, Brani Vidakovic, Felix Abramovich, and Theofanis Sapatinas. Optimal testing in functional analysis of variance models. Technical report, Georgia Institute of Technology, 2002.

[3] Peter J Brockwell and Richard A Davis. *Time series: theory and methods*. Springer science & business media, 2009.

[4] Stanley R Bull. Renewable energy today and tomorrow. *Proceedings of the IEEE*, 89 (8):1216–1226, 2001.

[5] Cody Carroll, Alvaro Gajardo, Yaqing Chen, Xiongtao Dai, Jianing Fan, Pantelis Z. Hadjipantelis, Kyunghee Han, Hao Ji, Hans-Georg Mueller, and Jane-Ling Wang. *fdapace: Functional Data Analysis and Empirical Dynamics*, 2021. URL https://CRAN.R-project.org/package=fdapace. R package version 0.5.6.

[6] George Casella and Roger L Berger. *Statistical inference*. Cengage Learning, 2021.

[7] Davide Castelvecchi et al. Electric cars and batteries: how will the world produce enough? *Nature*, 596(7872):336–339, 2021.

[8] Centers for Disease Control and Prevention. Chronic Kidney Disease in the United States, 2021. Technical report, US Department of Health and Human Services, Centers

for Disease Control and Prevention, Atlanta, GA, 2021. URL https://www.cdc.gov/kidneydisease/publications-resources/2019-national-facts.html.

[9] Antonio Cuevas, Manuel Febrero, and Ricardo Fraiman. An anova test for functional data. *Computational Statistics & Data Analysis*, 47(1):111 – 122, 2004. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2003.10.021. URL http://www.sciencedirect.com/science/article/pii/S016794730300269X.

[10] Boucar Diouf and Ramchandra Pode. Potential of lithium-ion batteries in renewable energy. *Renewable Energy*, 76:375–380, 2015.

[11] Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001. ISSN 01621459. URL http://www.jstor.org/stable/3085904.

[12] Jianqing Fan and Sheng-Kuei Lin. Test of significance when data are curves. *Journal of the American Statistical Association*, 93(443):1007–1021, 1998. doi: 10.1080/01621459.1998.10473763. URL https://doi.org/10.1080/01621459.1998.10473763.

[13] Kuangnan Fang, Xiaochen Zhang, Shuangge Ma, and Qingzhao Zhang. Smooth and locally sparse estimation for multiple-output functional linear regression. *Journal of Statistical Computation and Simulation*, 90(2):341–354, 2020. ISSN 15635163. doi: 10.1080/00949655.2019.1680676.

[14] Julian J. Faraway. Regression analysis for a functional response. *Technometrics*, 39(3):254–261, 1997. ISSN 15372723. doi: 10.1080/00401706.1997.10485118.

[15] Tomasz Górecki and Łukasz Smaga. A comparison of tests for the one-way ANOVA problem for functional data. *Computational Statistics*, 30(4):987–1010, 2015.

[16] Chong Gu. *Smoothing spline ANOVA models*, volume 297. Springer, 2013.

[17] Yi-Jun He, Jia-Ni Shen, Ji-Fu Shen, and Zi-Feng Ma. State of health estimation of lithium-ion batteries: A multiscale gaussian process regression modeling approach. *AIChE Journal*, 61(5):1589–1600, 2015.

[18] Lajos Horváth and Gregory Rice. Testing equality of means when the observations are from functional time series. *Journal of Time Series Analysis*, 36(1):84–108, 2015. ISSN 14679892. doi: 10.1111/jtsa.12095.

[19] Lajos Horváth, Piotr Kokoszka, and Ron Reeder. Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(1):103–122, 2013. doi: 10.1111/j.1467-9868.2012.01032.x. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2012.01032.x.

[20] Gareth M. James, Jing Wang, and Ji Zhu. Functional linear regression that's interpretable. *Annals of Statistics*, 37(5 A):2083–2108, 2009. ISSN 00905364. doi: 10.1214/08-AOS641.

[21] Cari G. Kaufman and Stephan R. Sainy. Bayesian functional ANOVA modeling using Gaussian process prior distributions. *Bayesian Analysis*, 5(1):123–150, 2010. ISSN 19360975. doi: 10.1214/10-BA505.

[22] Eloïse Lancelot, Philippe Courcoux, Sylvie Chevallier, Alain Le-Bail, and Benoît Jaillais. Prediction of water content in biscuits using near infrared hyperspectral imaging spectroscopy and chemometrics. *Journal of Near Infrared Spectroscopy*, 28(3):140–147, 2020.

[23] Yehua Li, Yumou Qiu, and Yuhang Xu. From multivariate to functional data analysis: Fundamentals, recent developments, and emerging areas. *Journal of Multivariate Analysis*, 188:104806, 2022. ISSN 0047-259X. doi: https://doi.org/10.1016/

j.jmva.2021.104806. URL https://www.sciencedirect.com/science/article/pii/S0047259X21000841. 50th Anniversary Jubilee Edition.

[24] Zhenhua Lin, Jiguo Cao, Liangliang Wang, and Haonan Wang. Locally sparse estimator for functional linear regression models. *Journal of Computational and Graphical Statistics*, 26(2):306–318, 2017. doi: 10.1080/10618600.2016.1195273. URL https://doi.org/10.1080/10618600.2016.1195273.

[25] Baisen Liu, Liangliang Wang, and Jiguo Cao. Estimating functional linear mixed-effects regression models. *Computational Statistics & Data Analysis*, 106:153–164, 2017.

[26] Chong Liu, Surajit Ray, and Giles Hooker. Functional principal component analysis of spatially correlated data. *Statistics and Computing*, 27(6):1639–1654, 2017.

[27] Datong Liu, Yue Luo, Yu Peng, Xiyuan Peng, and Michael Pecht. Lithium-ion battery remaining useful life estimation based on nonlinear ar model combined with degradation feature. In *Annual Conference of the PHM Society*, volume 4, 2012.

[28] Datong Liu, Jingyue Pang, Jianbao Zhou, and Yu Peng. Data-driven prognostics for lithium-ion battery based on gaussian process regression. In *Proceedings of the IEEE 2012 prognostics and system health management conference (PHM-2012 Beijing)*, pages 1–5. IEEE, 2012.

[29] Datong Liu, Jingyue Pang, Jianbao Zhou, Yu Peng, and Michael Pecht. Prognostics for state of health estimation of lithium-ion batteries based on combination gaussian process functional regression. *Microelectronics Reliability*, 53(6):832–839, 2013.

[30] Datong Liu, Yue Luo, Jie Liu, Yu Peng, Limeng Guo, and Michael Pecht. Lithium-ion battery remaining useful life estimation based on fusion nonlinear degradation ar model and rpf algorithm. *Neural Computing and Applications*, 25(3):557–572, 2014.

[31] Datong Liu, Jianbao Zhou, Dawei Pan, Yu Peng, and Xiyuan Peng. Lithium-ion battery remaining useful life estimation with an optimized relevance vector machine algorithm with incremental learning. *Measurement*, 63:143–151, 2015.

[32] Weilin Luo, Chao Lv, Lixin Wang, and Chao Liu. Study on impedance model of li-ion battery. In *2011 6th IEEE Conference on Industrial Electronics and Applications*, pages 1943–1947. IEEE, 2011.

[33] Jon Ander Martin, Justin N Ouwerkerk, Anthony P Lamping, and Kelly Cohen. Comparison of battery modeling regression methods for application to unmanned aerial vehicles. *Complex Engineering Systems*, 2022.

[34] William Meeker, Yili Hong, and Luis Escobar. Degradation models and analyses. *Encyclopedia of Statistical Sciences*, pages 1–23, 2004.

[35] Jeffrey S. Morris. Functional Regression. *Annual Review of Statistics and Its Application*, 2:321–359, 2015. doi: 10.1146/annurev-statistics-010814-020413.

[36] Renato G Nascimento, Matteo Corbetta, Chetan S Kulkarni, and Felipe AC Viana. Hybrid physics-informed neural networks for lithium-ion battery modeling and prognosis. *Journal of Power Sources*, 513:230526, 2021.

[37] Selina SY Ng, Yinjiao Xing, and Kwok L Tsui. A naive bayes model for robust remaining useful life prediction of lithium-ion battery. *Applied Energy*, 118:114–123, 2014.

[38] Xuanlong Nguyen and Alan E. Gelfand. Bayesian nonparametric modeling for functional analysis of variance. *Annals of the Institute of Statistical Mathematics*, 66(3):495–526, 2014. ISSN 15729052. doi: 10.1007/s10463-013-0436-7.

[39] NIH National Institute of Diabetes, Digestive, and Kidney Disease. Hemodialysis,

2018. URL https://www.niddk.nih.gov/health-information/kidney-disease/kidney-failure/hemodialysis.

[40] AG Olabi and Mohammad Ali Abdelkareem. Renewable energy and climate change. *Renewable and Sustainable Energy Reviews*, 158:112111, 2022.

[41] NL Panwar, SC Kaushik, and Surendra Kothari. Role of renewable energy sources in environmental protection: A review. *Renewable and sustainable energy reviews*, 15(3): 1513–1524, 2011.

[42] So Young Park and Ana-Maria Staicu. Longitudinal functional data analysis. *Stat*, 4 (1):212–226, 2015.

[43] Meru A Patil, Piyush Tagade, Krishnan S Hariharan, Subramanya M Kolake, Taewon Song, Taejung Yeo, and Seokgwang Doo. A novel multistage support vector machine based approach for li ion battery remaining useful life estimation. *Applied energy*, 159: 285–297, 2015.

[44] José Pinheiro, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, Siem Heisterkamp, Bert Van Willigen, and R Maintainer. Package 'nlme'. *Linear and nonlinear mixed effects models, version*, 3(1), 2017.

[45] Alessia Pini and Simone Vantini. The interval testing procedure: A general framework for inference in functional data analysis. *Biometrics*, 72(3):835–845, 2016. doi: https://doi.org/10.1111/biom.12476. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.12476.

[46] Alessia Pini, Simone Vantini, Bianca Maria Colosimo, and Marco Grasso. Domain-selective functional analysis of variance for supervised statistical profile monitoring

of signal data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 67(1):55–81, 2018. doi: https://doi.org/10.1111/rssc.12218. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssc.12218.

[47] JMP Pro. JMP functional data explorer. https://www.jmp.com/support/help/en/15.2/index.shtml#page/jmp/functional-data-explorer.shtml, 2021. Accessed: 2021-10-03.

[48] J. O. Ramsay, Spencer Graves, and Giles Hooker. *fda: Functional Data Analysis*, 2020. URL https://CRAN.R-project.org/package=fda. R package version 5.1.9.

[49] James O. Ramsay and Bernard W. Silverman. *Functional data analysis (1st Ed.)*. Springer Science Business Media, 1997.

[50] James O. Ramsay and Bernard W. Silverman. *Applied Functional Data Analysis: Methods and Case Studies*. Springer, 2002. ISBN 0387954147. doi: 10.1111/j.1467-985x.2004.t01-5-.x.

[51] James O. Ramsay and Bernard W. Silverman. *Functional data analysis (2nd Ed.)*. Springer Science Business Media, 2005.

[52] Sarah J Ratcliffe, Leo R Leader, and Gillian Z Heller. Functional data analysis with application to periodically stimulated foetal heart rate data. i: Functional regression. *Statistics in Medicine*, 21(8):1103–1114, 2002.

[53] Robert R Richardson, Michael A Osborne, and David A Howey. Gaussian process regression for forecasting battery state of health. *Journal of Power Sources*, 357:209–219, 2017.

[54] B Saha and K Goebel. Battery data set, nasa ames prognostics data repository

(http://ti. arc. nasa. gov/project/prognostic-data-repository). *NASA Ames Research Center, Moffett Field, CA. NASA AMES Prognostics Data Repository*, 2007.

[55] Bhaskar Saha and Kai Goebel. Modeling li-ion battery capacity depletion in a particle filtering framework. In *Annual Conference of the PHM Society*, volume 1, 2009.

[56] Bhaskar Saha, Kai Goebel, Scott Poll, and Jon Christophersen. Prognostics methods for battery health monitoring using a bayesian framework. *IEEE Transactions on instrumentation and measurement*, 58(2):291–296, 2008.

[57] Bhaskar Saha, Kai Goebel, and Jon Christophersen. Comparison of prognostic algorithms for estimating remaining useful life of batteries. *Transactions of the Institute of Measurement and Control*, 31(3-4):293–308, 2009.

[58] María Isabel Sánchez-Rodríguez, Elena M Sánchez-López, Alberto Marinas, Francisco José Urbano, and José M Caridad. Functional approach and agro-climatic information to improve the estimation of olive oil fatty acid content from near-infrared data. *Food Science & Nutrition*, 8(1):351–360, 2020.

[59] Claudio Sbarufatti, Matteo Corbetta, Marco Giglio, and Francesco Cadini. Adaptive prognosis of lithium-ion batteries based on the combination of particle filters and radial basis function neural networks. *Journal of Power Sources*, 344:128–140, 2017.

[60] Ryan S Senger, Varun Kavuru, Meaghan Sullivan, Austin Gouldin, Stephanie Lundgren, Kristen Merrifield, Caitlin Steen, Emily Baker, Tommy Vu, Ben Agnor, et al. Spectral characteristics of urine specimens from healthy human volunteers analyzed using raman chemometric urinalysis (rametrix). *PloS one*, 14(9):e0222115, 2019.

[61] Amrik Shah, Nan Laird, and David Schoenfeld. A random-effects model for multiple

characteristics with possibly missing data. *Journal of the American Statistical Association*, 92(438):775–779, 1997.

[62] Qing Shen and Julian Faraway. An f test for linear models with functional responses. *Statistica Sinica*, pages 1239–1257, 2004.

[63] Yan Shen, Lijuan Shen, and Wangtu Xu. A wiener-based degradation model with logistic distributed measurement errors and remaining useful life estimation. *Quality and Reliability Engineering International*, 34(6):1289–1303, 2018.

[64] Piyush Tagade, Krishnan S Hariharan, Sanoop Ramachandran, Ashish Khandelwal, Arunava Naha, Subramanya Mayya Kolake, and Seong Ho Han. Deep gaussian process regression for lithium-ion battery health prognosis and degradation mode diagnosis. *Journal of Power Sources*, 445:227281, 2020.

[65] Shengjin Tang, Chuanqiang Yu, Xue Wang, Xiaosong Guo, and Xiaosheng Si. Remaining useful life prediction of lithium-ion batteries based on the wiener process with measurement error. *energies*, 7(2):520–547, 2014.

[66] Rodolphe Thiébaut, Hélène Jacqmin-Gadda, Geneviève Chêne, Catherine Leport, and Daniel Commenges. Bivariate linear mixed models using sas proc mixed. *Computer methods and programs in biomedicine*, 69(3):249–256, 2002.

[67] United States Renal Data System. 2020 USRDS annual data report: Epidemiology of kidney disease in the United States. Technical report, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2020. URL https://adr.usrds.org/2020/end-stage-renal-disease/1-incidence-prevalence-patient-characteristics-and-treatment-modalities.

[68] Haonan Wang and Bo Kai. Functional sparsity: Global versus local. *Statistica Sinica*, 25

(4):1337–1354, 2015. ISSN 10170405, 19968507. URL http://www.jstor.org/stable/24721236.

[69] Jane-Ling Wang, Jeng-Min Chiou, and Hans-Georg Müller. Functional Data Analysis. *Annual Review of Statistics and Its Application*, 3(1):257–295, jun 2016. ISSN 2326-8298. doi: 10.1146/annurev-statistics-041715-033624. URL https://doi.org/10.1146/annurev-statistics-041715-033624.

[70] Yunnan Xu, Pang Du, John Robertson, and Ryan Senger. Sparse logistic regression on functional data. *Statistics and Its Interface*, 2021.

[71] Yunnan Xu, Pang Du, Ryan Senger, John Robertson, and James L Pirkle. ISREA: An efficient peak-preserving baseline correction algorithm for raman spectra. *Applied Spectroscopy*, 75(1):34–45, 2021.

[72] Fang Yao, Hans-Georg Müller, and Jane-Ling Wang. Functional data analysis for sparse longitudinal data. *Journal of the American statistical association*, 100(470):577–590, 2005.

[73] Zhilong Yu, Yekai Zhang, Lihua Qi, and Ran Li. Soh estimation method for lithium-ion battery based on discharge characteristics. *Int. J. Electrochem. Sci*, 17(220725):2, 2022.

[74] Jin-Ting Zhang. Statistical inferences for linear models with functional responses. *Statistica Sinica*, pages 1431–1451, 2011.

[75] Jin-Ting Zhang. *Analysis of variance for functional data*. CRC Press, 2013.

[76] Jin Ting Zhang and Xuehua Liang. One-way ANOVA for functional data via globalizing the pointwise F-test. *Scandinavian Journal of Statistics*, 41(1):51–71, 2014. ISSN 03036898. doi: 10.1111/sjos.12025.

[77] Jin-Ting Zhang, Jianwei Chen, et al. Statistical inferences for functional data. *The Annals of Statistics*, 35(3):1052–1079, 2007.

[78] Jin-Ting Zhang, Ming-Yen Cheng, Hau-Tieng Wu, and Bu Zhou. A new test for functional one-way ANOVA with applications to ischemic heart screening. *Computational Statistics & Data Analysis*, 132:3 – 17, 2019. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2018.05.004. URL http://www.sciencedirect.com/science/article/pii/S0167947318301130. Special Issue on Biostatistics.

[79] Jianhui Zhou, Nae Yuh Wang, and Naisyin Wang. Functional linear model with zero-value coefficient function at sub-regions. *Statistica Sinica*, 23(1):25–50, 2013. ISSN 10170405. doi: 10.5705/ss.2010.237.

Appendices

Appendix A

Additional Results from Simulation Study of Chapter 2

A.1 Results for settings with sample size $N = (10, 20, 30)$

A.1.1 Setting 1: True contrast is a common-form smooth function

Errors come from Brownian Bridge process

	Test	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
Common group-wise covariance function	L2N	5	10	60	100	100	100	100	100
	L2B	6	11	64	100	100	100	100	100
	L2b	8	13	72	99	100	100	100	100
	FN	4	9	57	100	100	100	100	100
	FB	5	10	59	100	100	100	100	100
	Fb	19	49	99	100	100	100	100	100
	FT2-85	9	26	54	85	91	93	95	95
	FT2-95	21	67	96	100	100	100	100	100
Different group-wise covariance functions	L2N-het	4	8	41	95	100	100	100	100
	L2B-het	11	23	78	99	100	100	100	100
	L2b-het	4	8	46	95	100	100	100	100
	FN-het	3	7	37	88	100	100	100	100
	FB-het	4	7	42	90	100	100	100	100
	Fb-het	3	5	28	83	99	100	100	100
	FT2-85	5	24	56	84	89	99	99	99
	FT2-95	13	69	96	100	100	100	100	100

Table A.1: Empirical rejection percentages: the common-form contrast function setting with Brownian Bridge errors for $N = (10, 20, 30)$ setting by a common or different group-wise covariance functions.

Errors come from FAR(1) process

	Test	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
Common	L2N	5	12	72	100	100	100	100	100
group-wise	L2B	5	14	73	100	100	100	100	100
covariance	L2b	9	22	71	96	100	100	100	100
function	FN	5	12	70	100	100	100	100	100
	FB	5	12	71	100	100	100	100	100
	Fb	28	53	93	100	100	100	100	100
	FT2-85	8	31	58	72	81	85	88	89
	FT2-95	21	63	91	97	100	100	100	100
Different	L2N-het	13	21	56	90	100	100	100	100
group-wise	L2B-het	28	39	79	98	100	100	100	100
covariance	L2b-het	15	21	58	90	100	100	100	100
functions	FN-het	13	20	49	87	100	100	100	100
	FB-het	13	20	54	89	100	100	100	100
	Fb-het	13	15	40	84	100	100	100	100
	FT2-85	21	27	50	72	85	94	97	98
	FT2-95	25	59	98	99	100	100	100	100

Table A.2: Empirical rejection percentages: the common-form contrast function setting with FAR(1) errors for $N = (10, 20, 30)$ setting by a common or different group-wise covariance functions.

A.1.2 Setting 2: True contrast function is a spectrum-like function

Errors come from Brownian Bridge process

	Test	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
Common	L2N	6	8	49	100	100	100	100	100
group-wise	L2B	6	9	56	100	100	100	100	100
covariance	L2b	6	10	55	95	100	100	100	100
function	FN	5	8	45	100	100	100	100	100
	FB	6	8	49	100	100	100	100	100
	Fb	22	40	93	100	100	100	100	100
	FT2-85	18	19	22	26	32	44	59	63
	FT2-95	33	37	56	70	79	86	91	93
Different	L2N-het	4	8	37	90	100	100	100	100
group-wise	L2B-het	16	24	73	99	100	100	100	100
covariance	L2b-het	5	8	40	92	100	100	100	100
functions	FN-het	4	8	32	88	100	100	100	100
	FB-het	4	8	36	89	100	100	100	100
	Fb-het	4	6	26	83	99	100	100	100
	FT2-85	13	16	21	29	36	50	53	58
	FT2-95	25	34	64	78	88	92	93	96

Table A.3: Empirical rejection percentages: the spectrum-like contrast function setting with Brownian Bridge errors for $N = (10, 20, 30)$ setting by a common or different group-wise covariance functions.

Errors come from FAR(1) process

		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
Common	L2N	11	16	51	100	100	100	100	100
group-wise	L2B	11	16	52	100	100	100	100	100
covariance	L2b	13	21	62	98	100	100	100	100
function	FN	11	16	48	100	100	100	100	100
	FB	11	16	49	100	100	100	100	100
	Fb	34	49	98	100	100	100	100	100
	FT2-85	15	16	21	29	37	43	49	53
	FT2-95	29	31	50	63	77	81	87	89
Different	L2N-het	18	22	45	97	100	100	100	100
group-wise	L2B-het	30	38	78	100	100	100	100	100
covariance	L2b-het	19	23	46	98	100	100	100	100
functions	FN-het	18	21	41	93	100	100	100	100
	FB-het	18	21	45	94	100	100	100	100
	Fb-het	16	18	36	82	100	100	100	100
	FT2-85	16	15	17	25	29	40	44	52
	FT2-95	20	28	41	63	78	86	89	92

Table A.4: Empirical rejection percentages: the spectrum-like contrast function setting with FAR(1) errors for $N = (10, 20, 30)$ setting by a common or different group-wise covariance functions.

A.2 Additional simulation study to examine size control

A.2.1 Brownian bridge errors

	N	(10,10,10)	(10,20,30)	(50,50,50)	(100,100,100)	(200,200,200)
Common	L2N	5.2	4.4	4.6	4.8	4.4
group-wise	L2B	5.8	5.0	4.6	5.0	4.4
covariance	L2b	5.4	6.2	4.4	4.8	4.2
function	FN	5.0	4.0	4.4	4.8	4.4
	FB	5.2	4.2	4.6	4.8	4.4
	Fb	17.6	17.4	20.6	21.4	22.6
	FT2-95	21.6	19.4	8.2	6.0	5.2
Different	L2N-het	5.2	5.4	4.6	6.8	6.0
group-wise	L2B-het	15.6	11.6	12.0	12.6	14.2
covariance	L2b-het	5.8	5.4	4.2	6.8	6.0
functions	FN-het	4.6	4.6	4.4	6.8	6.0
	FB-het	5.0	5.2	4.4	6.8	6.0
	Fb-het	3.2	3.8	3.6	6.6	6.0
	FT2-95	20.4	17.6	9.4	5.8	5.0

Table A.5: Empirical Type I errors in percentage for common-form contrast function setting with Brownian Bridge errors using 500 replications

A.2.2 FAR(1) errors

	N	(10,10,10)	(10,20,30)	(50,50,50)	(100,100,100)	(200,200,200)
Common	L2N	11.8	9.8	11.8	12.2	11.4
group-wise	L2B	13.0	10.2	12.0	12.4	11.4
covariance	L2b	12.8	13.0	11.8	12.2	11.2
functions	FN	11.0	9.4	11.8	12.0	11.4
	FB	11.2	9.8	11.8	12.2	11.4
	Fb	28.4	30.0	33.2	32.2	30.4
	FT2-95	21.8	21.4	10.2	8.2	9.0
Different	L2N-het	14.0	12.2	10.2	13.0	11.0
group-wise	L2B-het	25.0	22.6	21.4	23.0	21.0
covariance	L2b-het	15.2	13.2	10.2	13.2	10.8
functions	FN-het	12.2	11.2	9.8	13.0	10.8
	FB-het	13.6	12.2	10.0	13.0	10.6
	Fb-het	10.6	10.6	9.4	13.0	11.6
	FT2-95	22.0	21.6	11.2	8.2	7.4

Table A.6: Empirical Type I errors for common-form contrast function setting with FAR(1) errors using 500 replications

Appendix B

Additional Plots from Simulation Study of Chapter 3

B.1 Simulation setting 1: Data from common functions

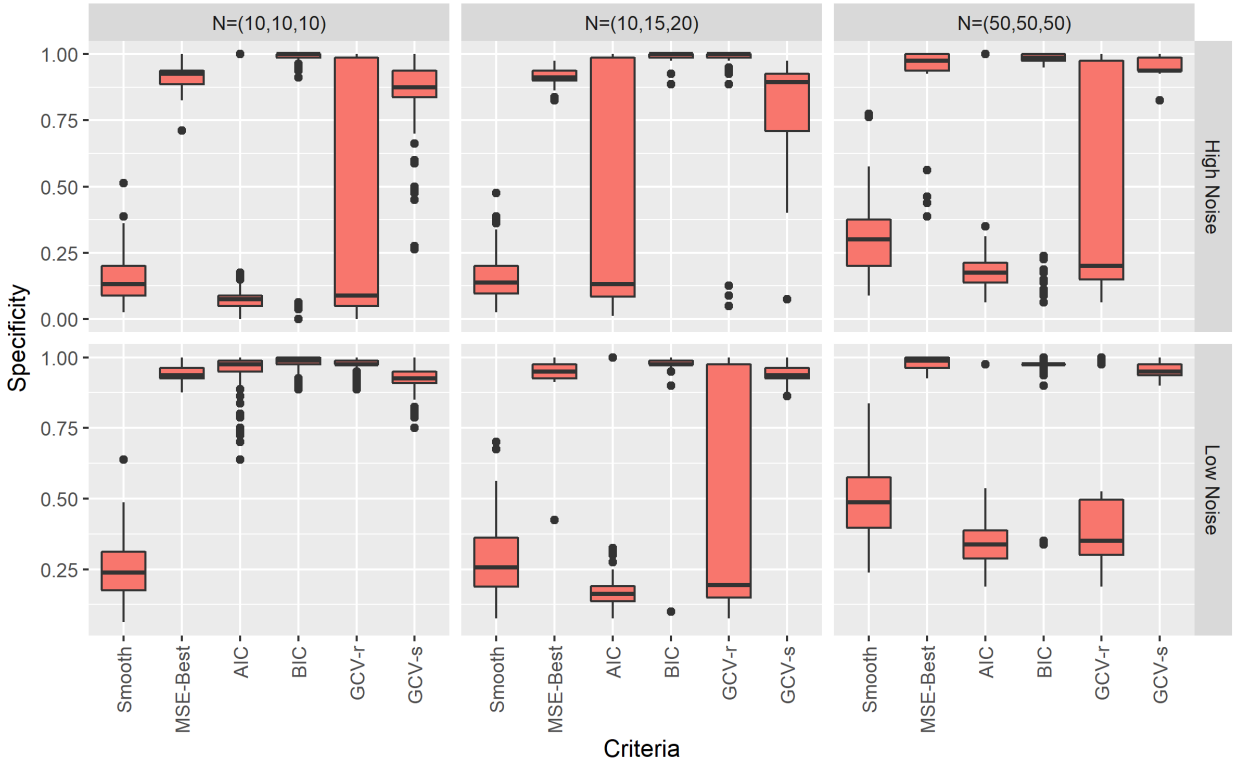


Figure B.1: Common function case: Specificity of the smoothed and sparsed contrast function from Gaussian noise case using 7 selection criteria

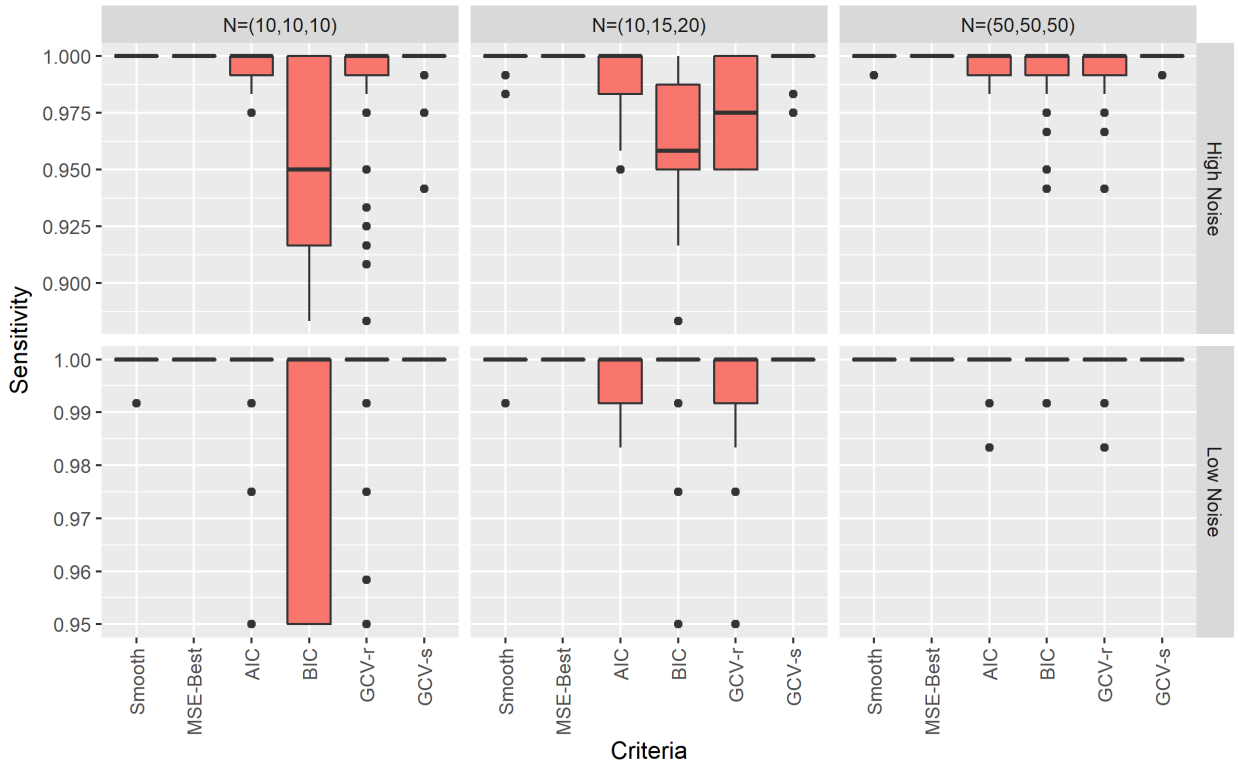


Figure B.2: Common function case: Sensitivity of the smoothed and sparsed contrast function from Gaussian noise case using 7 selection criteria

B.2 Simulation setting 2: Spectral data as observed data

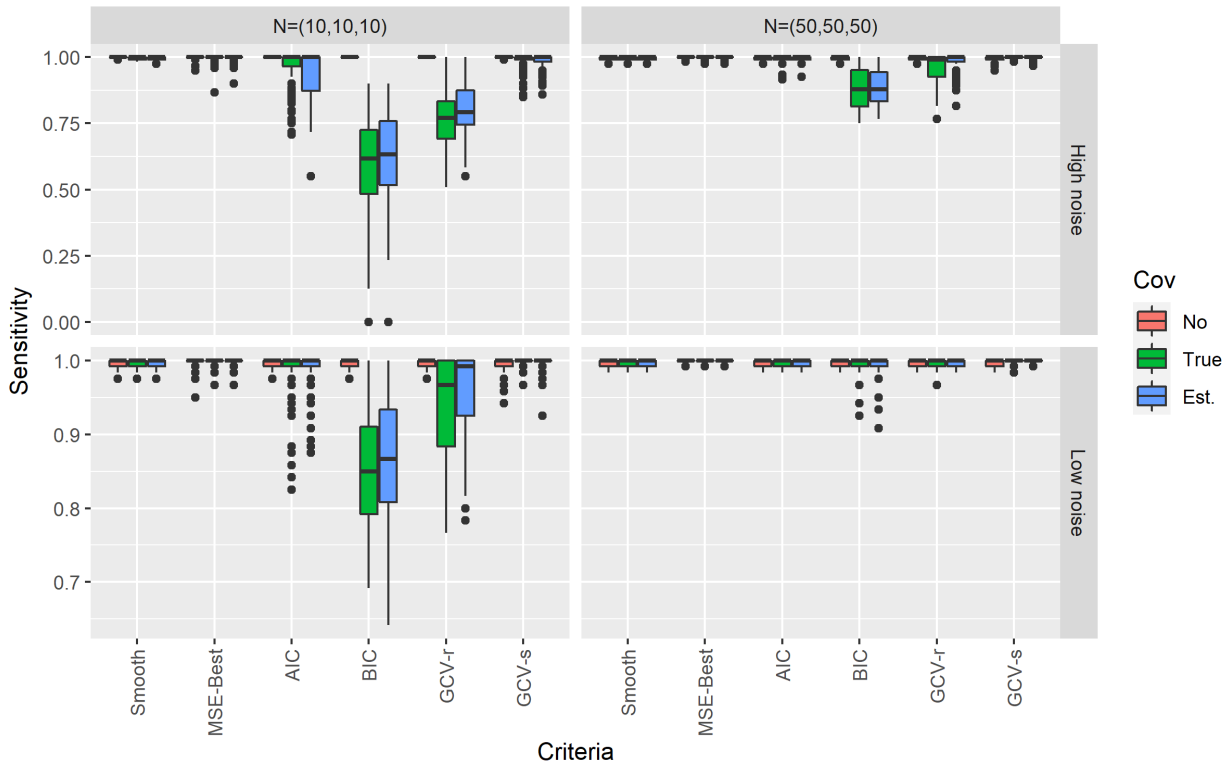


Figure B.3: Common function case: Sensitivity of the smoothed and sparsed contrast function from AR(1) noise case using 7 selection criteria

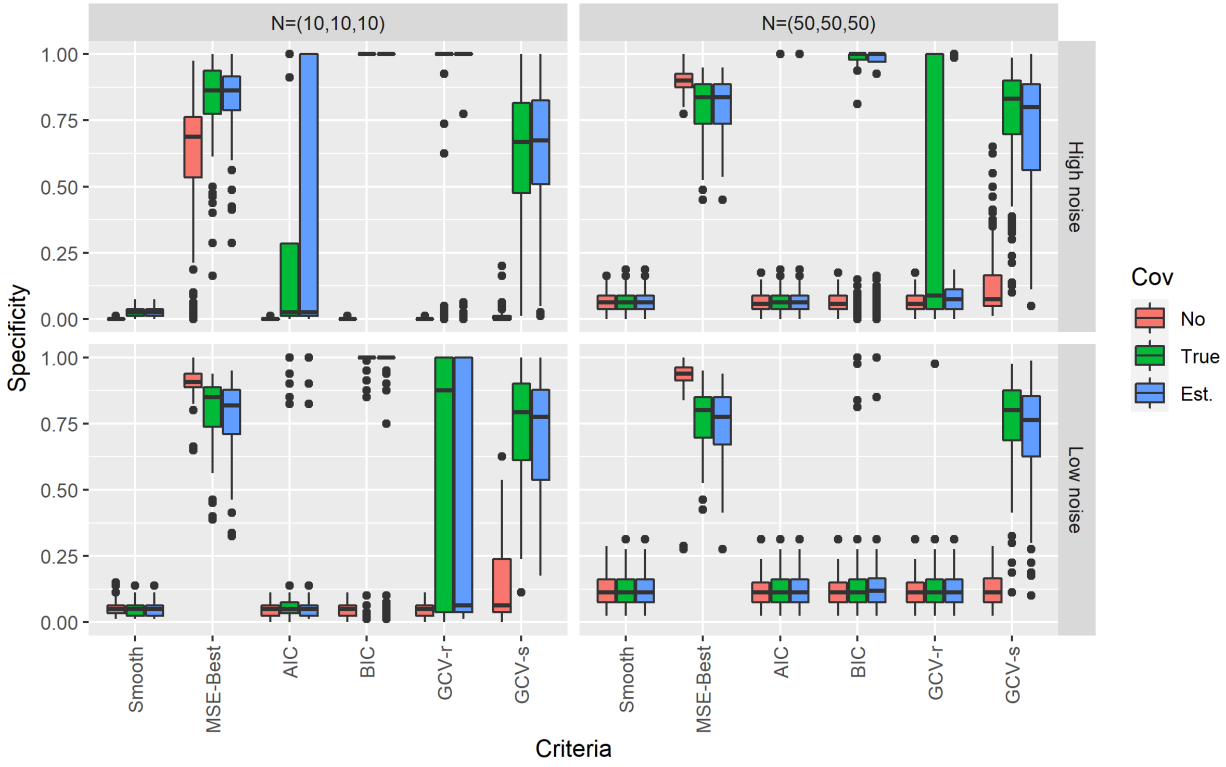


Figure B.4: Common function case: Specificity of the smoothed and sparsed contrast function from AR(1) noise case using 7 selection criteria

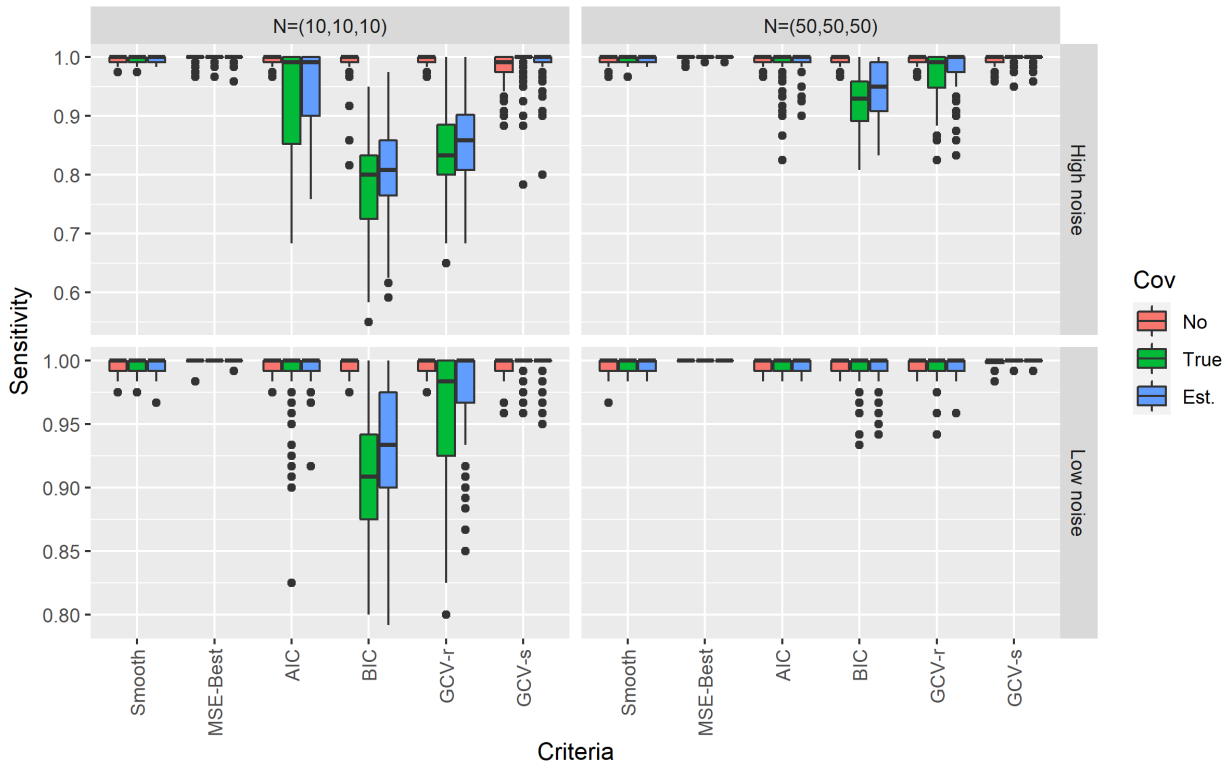


Figure B.5: Common function case: Sensitivity of the smoothed and sparsed contrast function from AR(7) noise case using 7 selection criteria

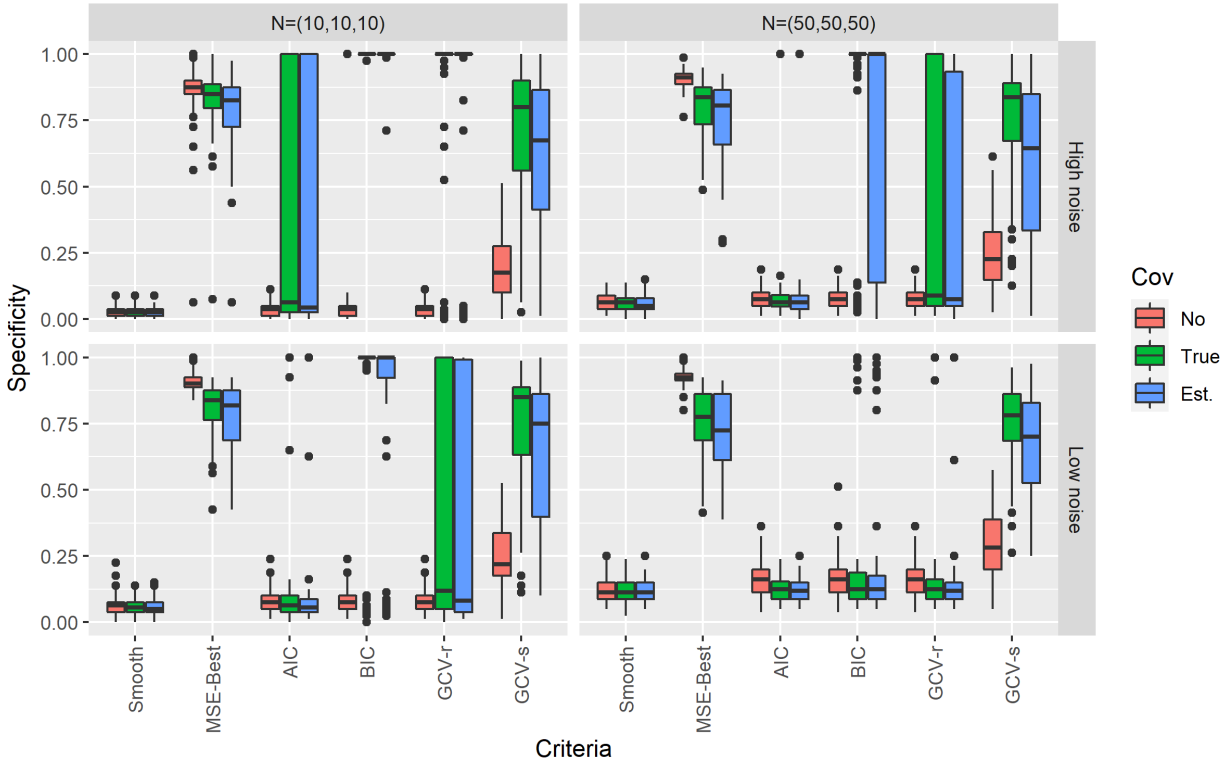


Figure B.6: Common function case: Specificity of the smoothed and sparsed contrast function from AR(7) noise case using 7 selection criteria

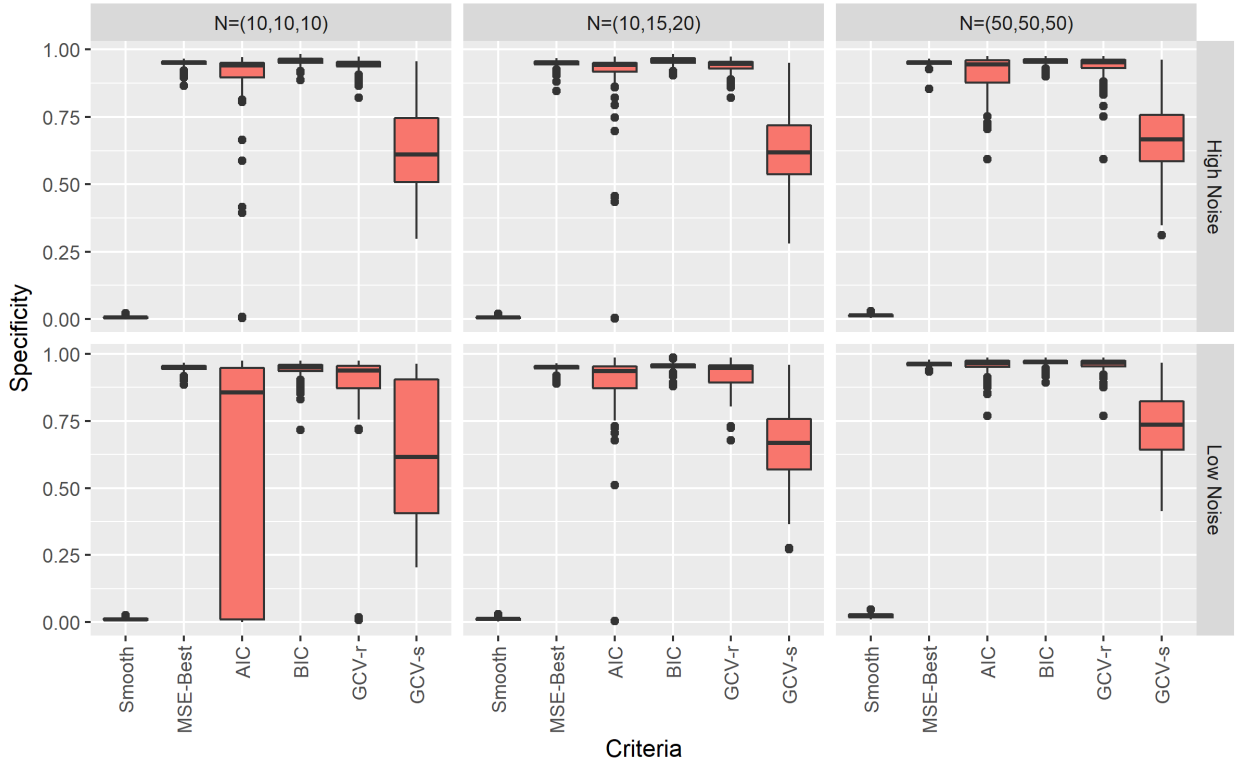


Figure B.7: Spectral data case: Specificity of the smoothed and sparsed contrast function from Gaussian noise case using 7 selection criteria

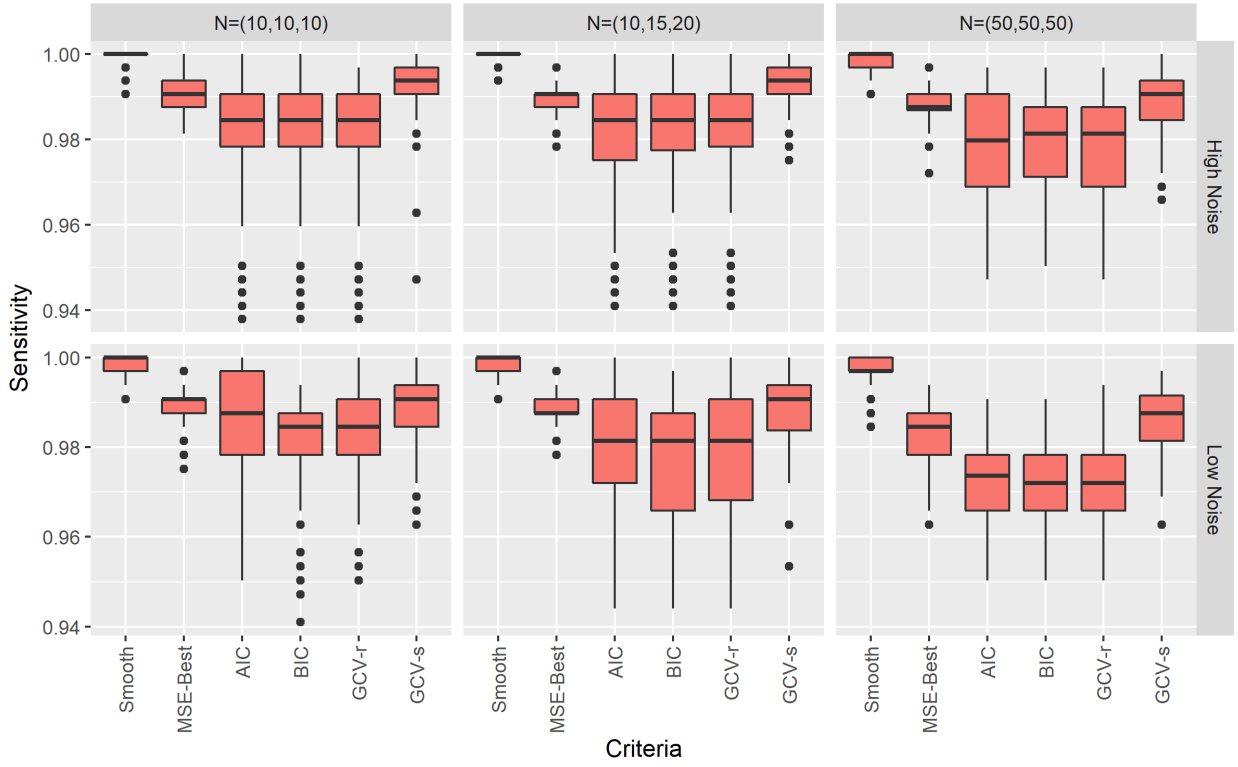


Figure B.8: Spectral data case: Sensitivity of the smoothed and sparsed contrast function from Gaussian noise case using 7 selection criteria

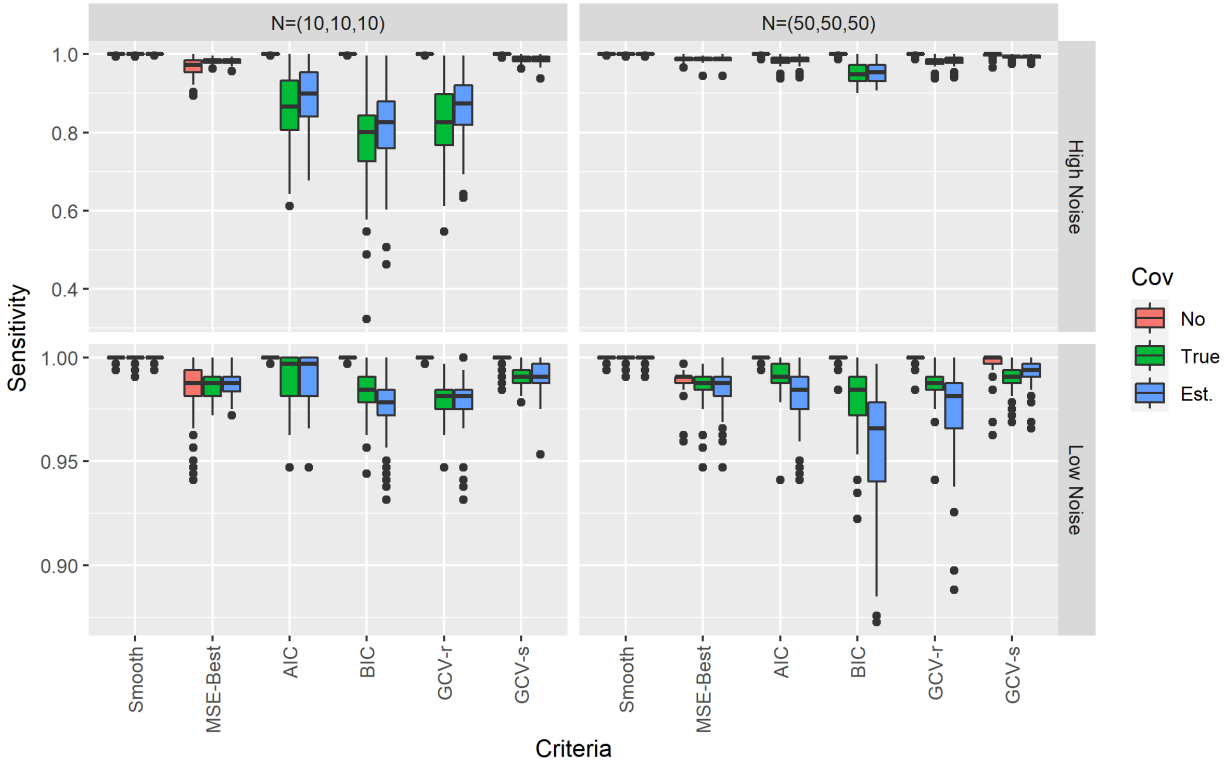


Figure B.9: Spectral data case: Sensitivity of the smoothed and sparsed contrast function from AR(1) noise case using 7 selection criteria

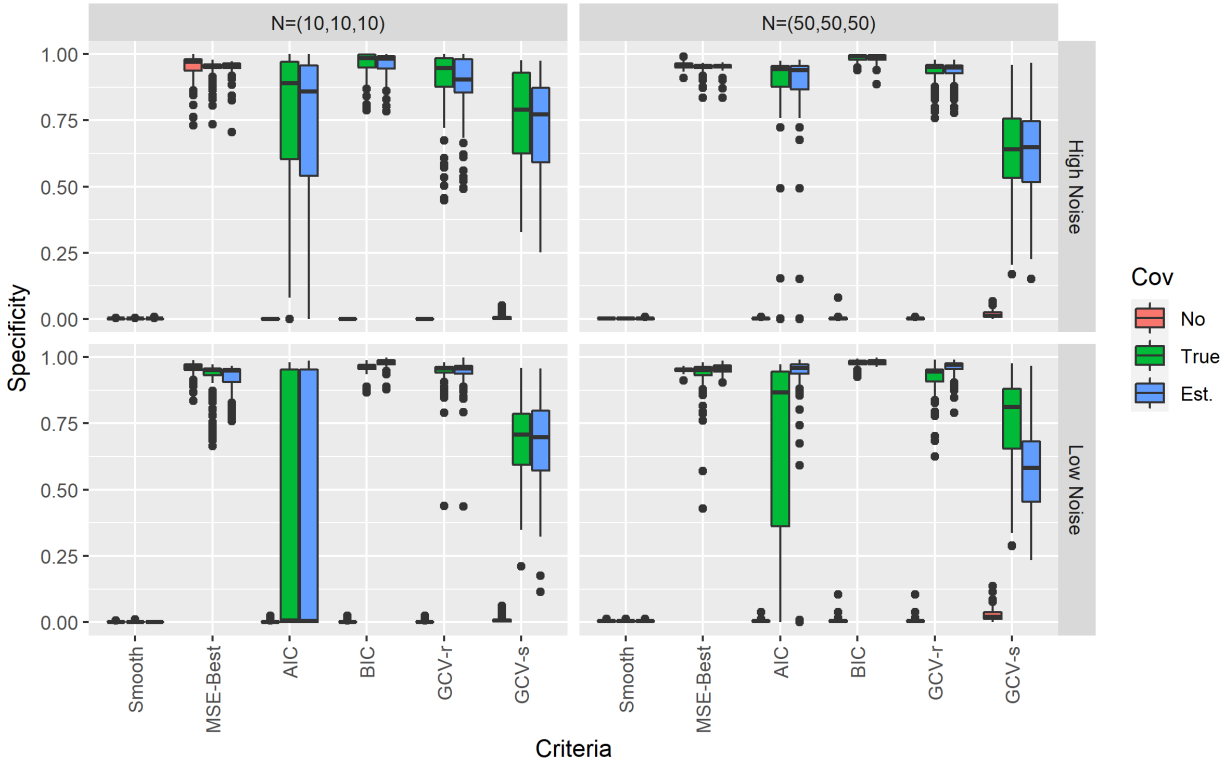


Figure B.10: Spectral data case: Specificity of the smoothed and sparsed contrast function from AR(1) noise case using 7 selection criteria

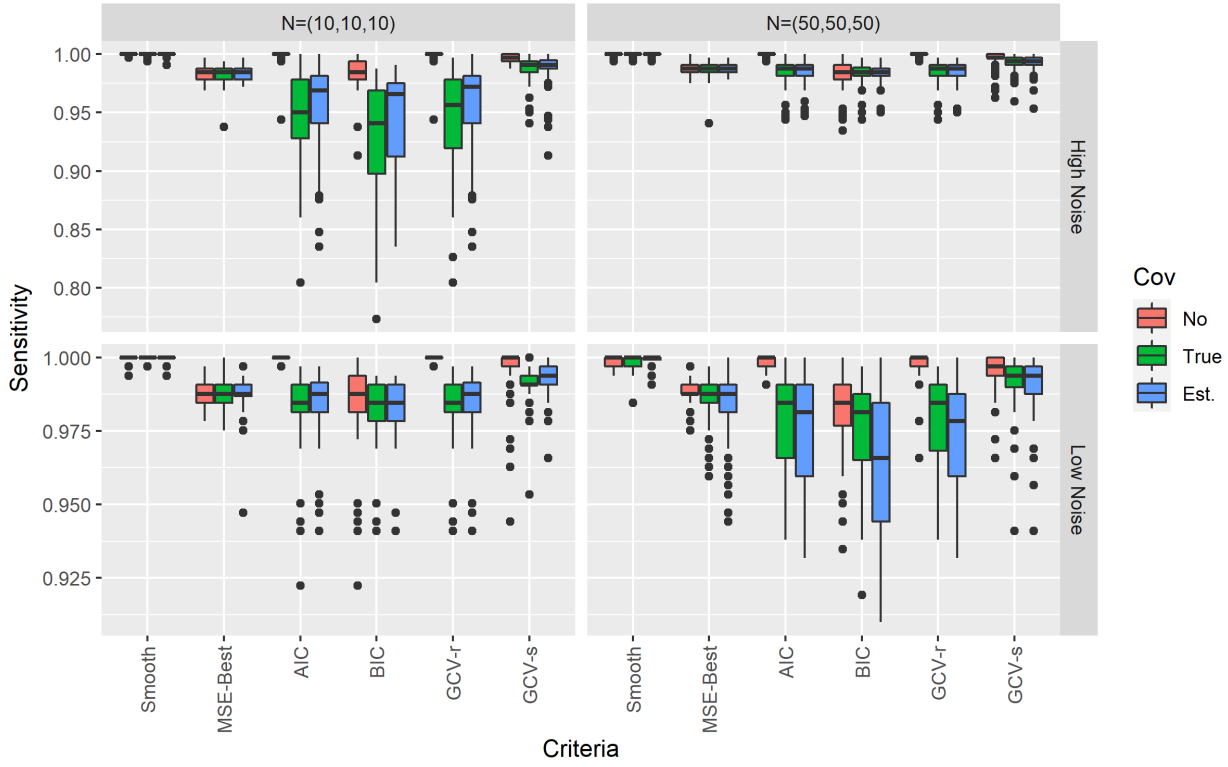


Figure B.11: Spectral data case: Sensitivity of the smoothed and sparsed contrast function from AR(7) noise case using 7 selection criteria

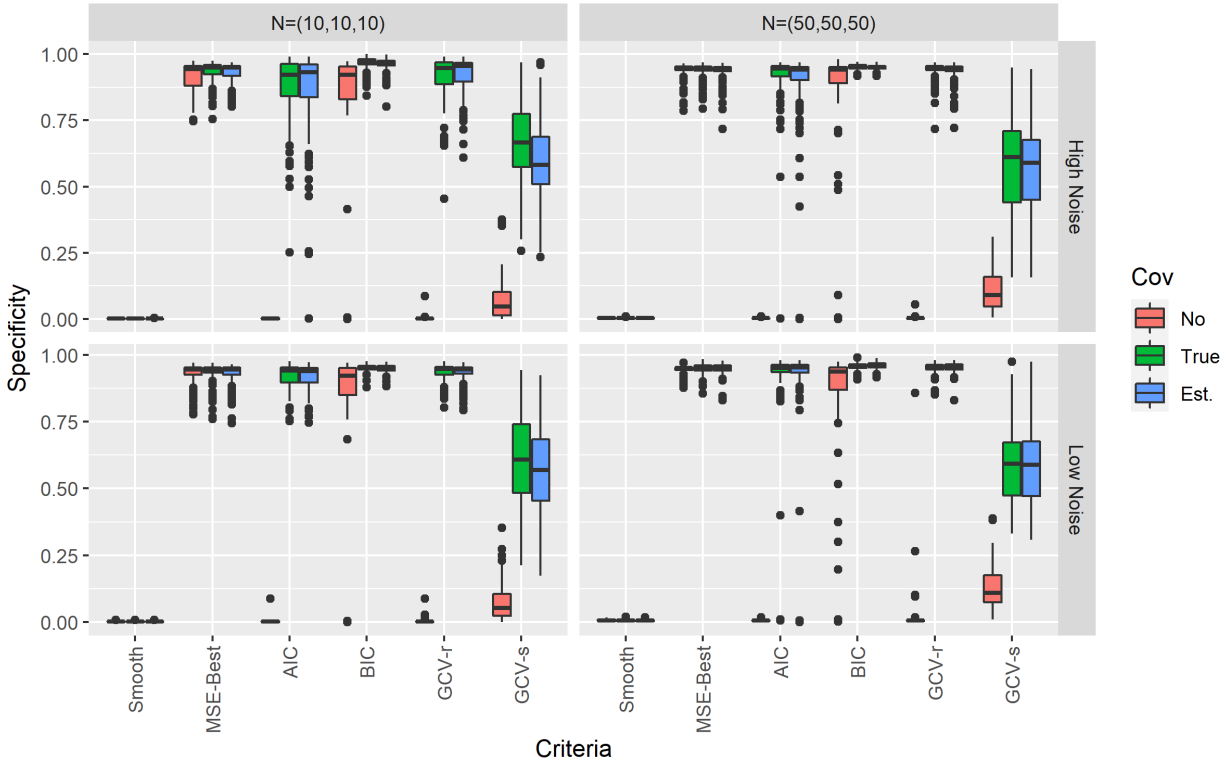


Figure B.12: Spectral data case: Specificity of the smoothed and sparsed contrast function from AR(7) noise case using 7 selection criteria

Appendix C

Additional Materials from Data Analysis of Chapter 4

C.1 Interpretation of functional principal components for scaled discharge voltage curves

Figure C.1 visualizes the first three estimated functional principal components obtained from FPC decompositon of the scaled VDCs from all available discharge data. This decomposition result is used in Section 4.5.3. Visualizing the FPCs by adding and subtracting them from the mean curves helps illustrate their interpretation. The first FPC, accounting for 89.54% of total variation in the data, is a positive function across the domain. Besides having values tapering toward 0 at the two ends, the first FPC is quite flat and its values consistent throughout the domain. The top right plot in Figure C.1 suggests that scaled VDC with positive score for the first FPC have higher voltage than the average scaled VDC almost entirely throughout the domain or discharge cycle. On the other hand, scaled VDC with negative score has lower voltage level than the average scaled VDC on the entire domain. Hence, first FPC indicates the deviation from the mean scaled VDC. The sign and magnitude of the first FPC score quantify the direction and strength of this deviation.

The second FPC explains 8.4% of total variation of the data. It has positive values approx-

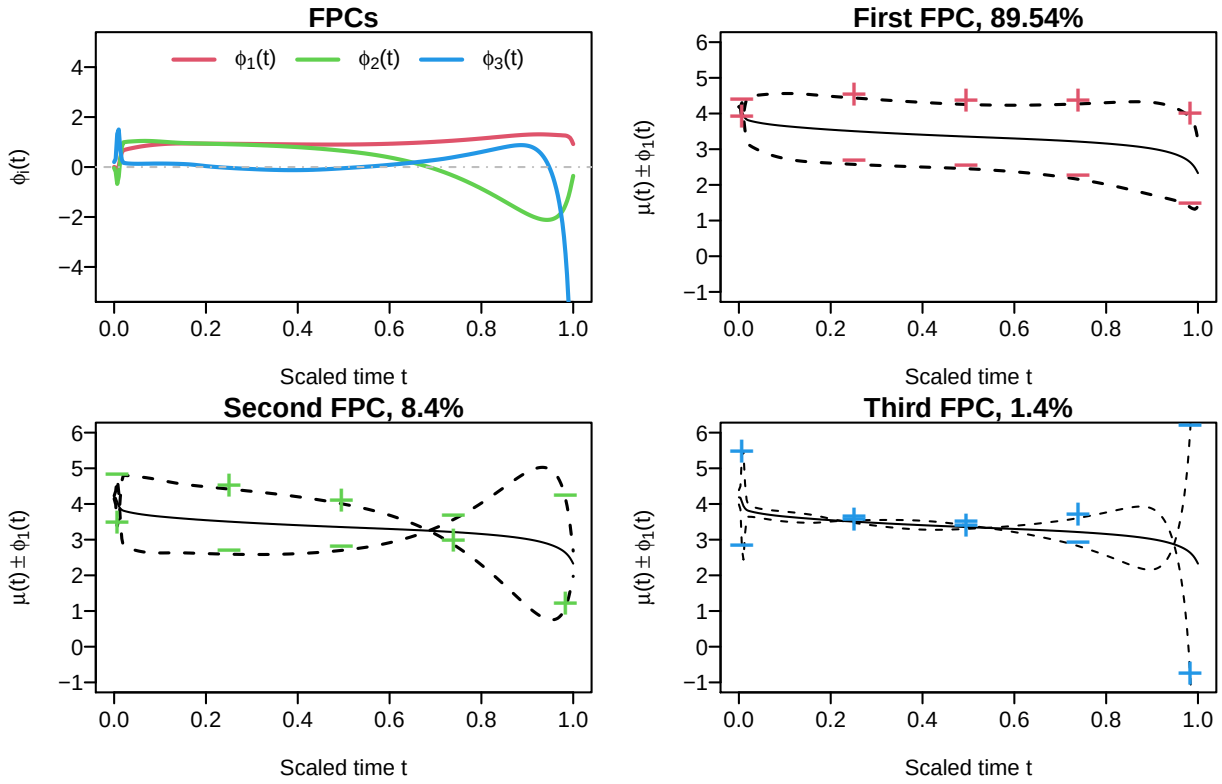


Figure C.1: First three principal component functions with their proportion of total variation explained from FPC decomposition of scaled VDCs for available data from 20 batteries. Top left show the plots of the three principal components. Top right and the bottom two plots show the effect on the mean function (solid black) after adding and subtracting the FPCs.

imately before 0.7 mark and negative after that. It represents the difference in discharge voltage in the first 70% of the cycle. This second FPC represents the contrast between voltage level in the first 70% of the discharge cycle and the latter 30%. Scaled VDCs with high second FPC scores are those with higher than average voltage in the first 70% of the discharge cycle but the voltage levels drop more significantly than the average in the remaining time of the cycle. Cycles with negative scores are those with lower voltage at the beginning of the cycle but the drop in later into the cycle is not as dramatic compared to the mean curve.

The third FPC shows close to 0 values for the majority of the cycle except for the beginning

and end of the cycle. This represents the variation in the scaled VDCs in terms of the beginning voltage level and how voltage drop at the end of the discharge cycles.

C.2 Interpretation of estimates from functional linear mixed effects model for EOD time

The model for end of discharge time b_{ic} applied in Section 4.5.3 is

$$b_{ic} = \alpha_0 + w_{0i} + (\alpha_1 + w_{1i})c + \int_0^1 (\beta(t) + b_i(t))\, x_{ic}(t)dt + \alpha_2 b_{i,c-1} + \alpha_3 \exp(-1/\Delta_{ic}) + \epsilon_{ic}, \quad \text{(C.1)}$$

The estimations for fixed coefficients and their standard errors (se) are $\hat{\alpha}_0 = -3845.733$ with standard error 931.087, $\hat{\alpha}_1 = -1.687$ with standard error 0.38, $\hat{\alpha}_2 = 546.7246$ with standard error 25.2, and $\hat{\alpha}_3 = 0.59$ with standard error 0.014. Figure C.2 illustrates the estimated coefficient function $\beta(t)$ of the effect of scaled VDCs on EOD plotted point-wise and Figure C.3 shows the 20 battery-specific effect from scaled VDCs on EOD, $\beta(t) + b_i(t)$.

From the form of (C.1), the effect from the scaled VDCs is a weighted average of the scaled VDC with the weight set by the coefficient function $\beta(t)$ adjusted with $b_i(t)$ for each battery. The estimated $\beta(t)$ has the shape close to a straight line with negative slope and sign of values switch at around 0.7. Recall the shape of a typical scaled VDCs (Figure C.1), we can see that this term $\int_0^1 (\beta(t) + b_i(t))\, x_{ic}(t)dt$ acts as an adjustment to EOD based on a weighted average of scaled VDCs. Without battery specific effect $b_i(t)$, this adjustment is similar to an area under the curve. Depending on the battery, the variations in the shapes of scaled VDCs help set the weight to be $\beta(t) + b_i(t)$.

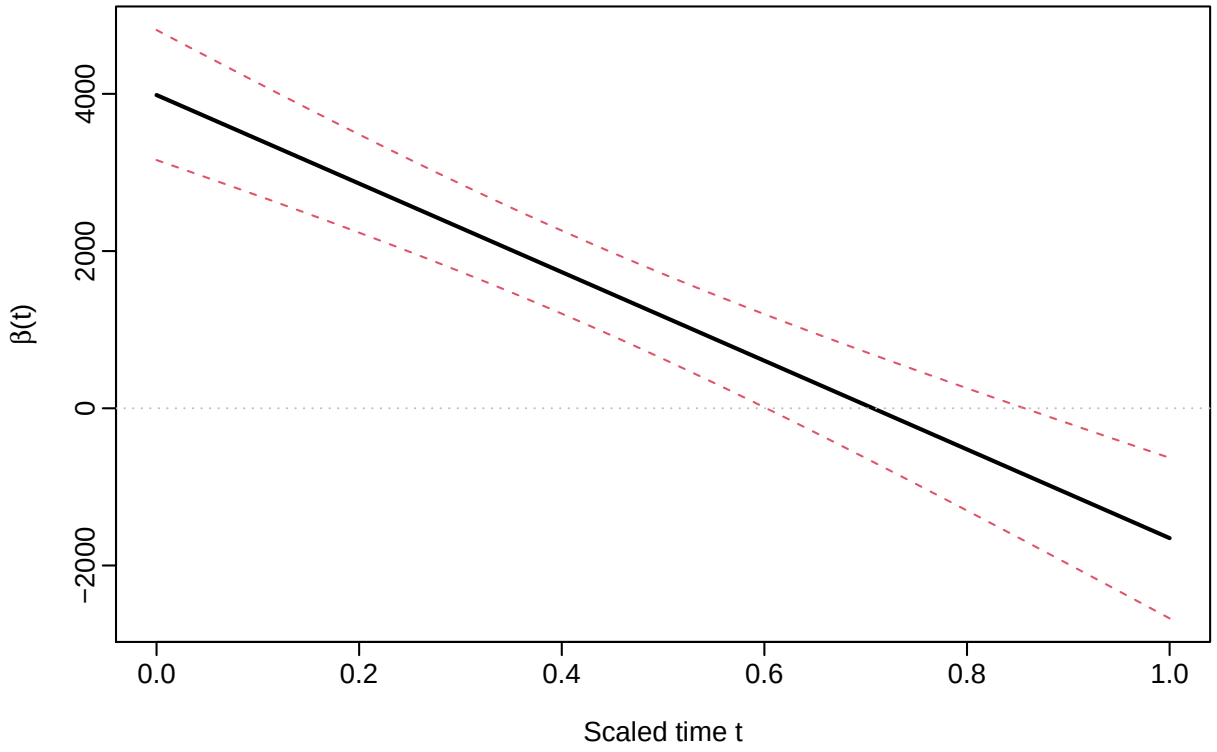


Figure C.2: Point-wise estimated coefficient function of the effect from scaled VDC on EOD time, $\beta(t)$. Mean estimate is in the solid black curve and red dashed curves indicated the approximated 95% confidence band.

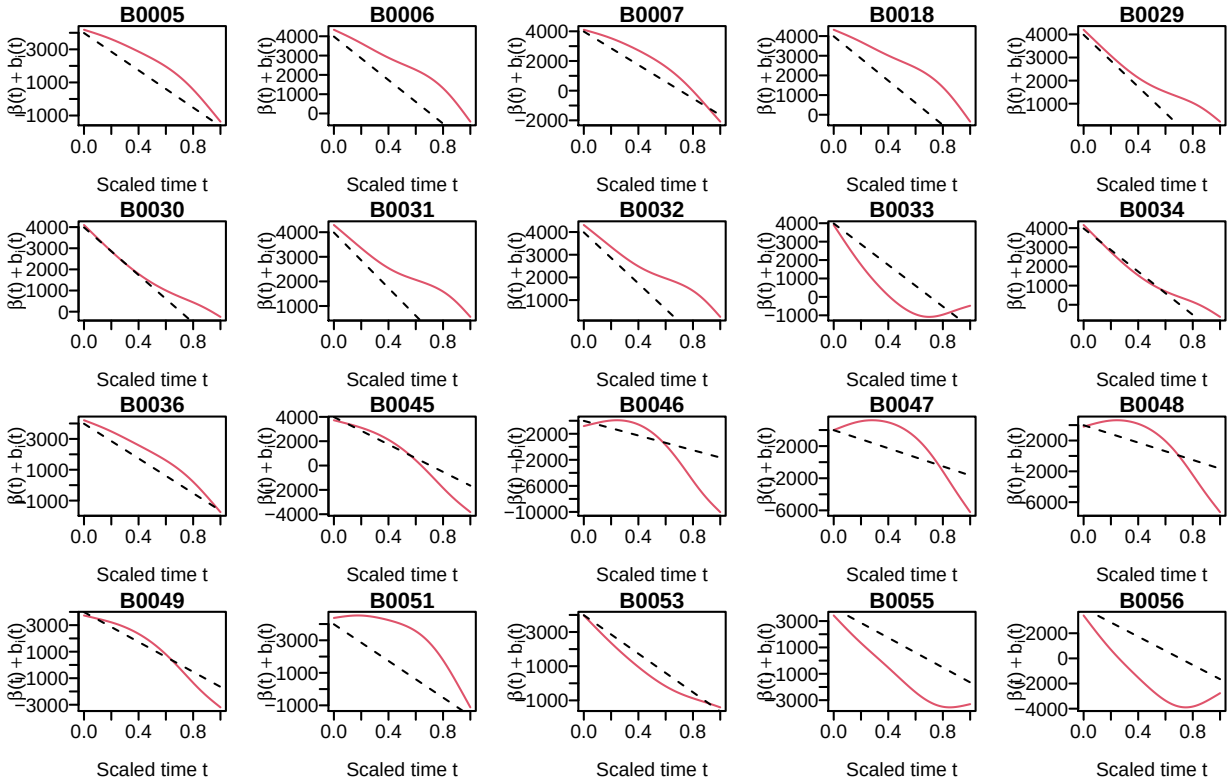


Figure C.3: Estimated functional coefficients of scaled VDC by battery from FLMM model predicting EOD time. Dashed curve represents overall coefficient function $\hat{\beta}(t)$. Battery-specific coefficient functions, $\hat{\beta}(t) + \hat{b}_i(t)$, are depicted by solid red curves